AD_____

Award Number: W81XWH-05-1-0267

TITLE: Functional Proteomic Analysis of Signaling Networks and Response to Targeted Therapy

PRINCIPAL INVESTIGATOR: Prahlad T Ram, Ph.D.

CONTRACTING ORGANIZATION: University of Texas
MD Anderson Cancer Center
Houston, Texas 77030

REPORT DATE: March 2008

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
01-03-2008	Annual	21 Feb 2007 – 20 Feb 2008

4. TITLE AND SUBTITLE

Functional Proteomic Analysis of Signaling Networks and Response to Targeted Therapy

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-05-1-0267

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Prahlad T Ram, Ph.D.

E-Mail: pram@mdanderson.org

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Texas
MD Anderson Cancer Center
Houston, Texas 77030

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The purpose of the research done has been to determine the regulation of the EGFR network and identify how manipulations of the network alter signal flow to bypass targeted inhibitions. The scope of the project is to understand the network and determine which molecules have to be targeted to inhibit tumor cell proliferation. The major finding thus far are 1) We have performed the proteomic analysis of the signaling network in a panel of 4 breast cancer cell lines and determined the network response to EGF and targeted inhibitors. 2) We have determined how information flows within the network and feedback regulation. 3) Using these biological data we have developed a computational model and have made predictions to identify combinations of targets. 4) We have experimentally tested the predictions on two cell lines with different mutations. We have completed the tasks listed for this time period and are on track as indicated in the original grant.

15. SUBJECT TERMS
No subject terms provided

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRMC
U	U	U	UU	64	**19b. TELEPHONE NUMBER** *(include area code)*

Table of Contents

Page

Prahlad T. Ram
Functional proteomic analysis of signaling networks and response to targeted therapy
DOD-IDEA BC044268

Progress report year 3

Introduction

The purpose of the research done has been to determine the regulation of the EGFR network and identify how manipulations of the network alter signal flow to bypass targeted inhibitions. The scope of the project is to understand the network and determine which molecules have to be targeted to inhibit tumor cell proliferation. In the past year we have been very active in our research efforts. We have accomplished many of the tasks laid out in the SOW. Task 1A, 1B, 1C, 1D, and 2A were completed during years 1 and 2. We have been working on Tasks 2B and 2A during the past funding year and the data from these two tasks are shown in this annual update.

Body

During this year we completed the first two aims of the proposal and expanded on these aims to include targeted inhibitors in addition to the siRNA in the initial proposal. The reason for including the targeted pharmacological inhibitors was the clinical relevance as these drugs are in early stage clinical trails. We used an AKT inhibitor, MEK inhibitor and two EGFR kinase inhibitors. We determined the phospho-proteomic profiles after pharmacological treatment in the panel of 6 breast tumor cell lines (MDA-231, BT549, MCF10A, T47D, MDA488, SKBR3). We discovered an unexpected finding in that MEK inhibitors were increase AKT phosphorylation in these cell lines. This was an important finding which we have been following in addition to the goals of the project.

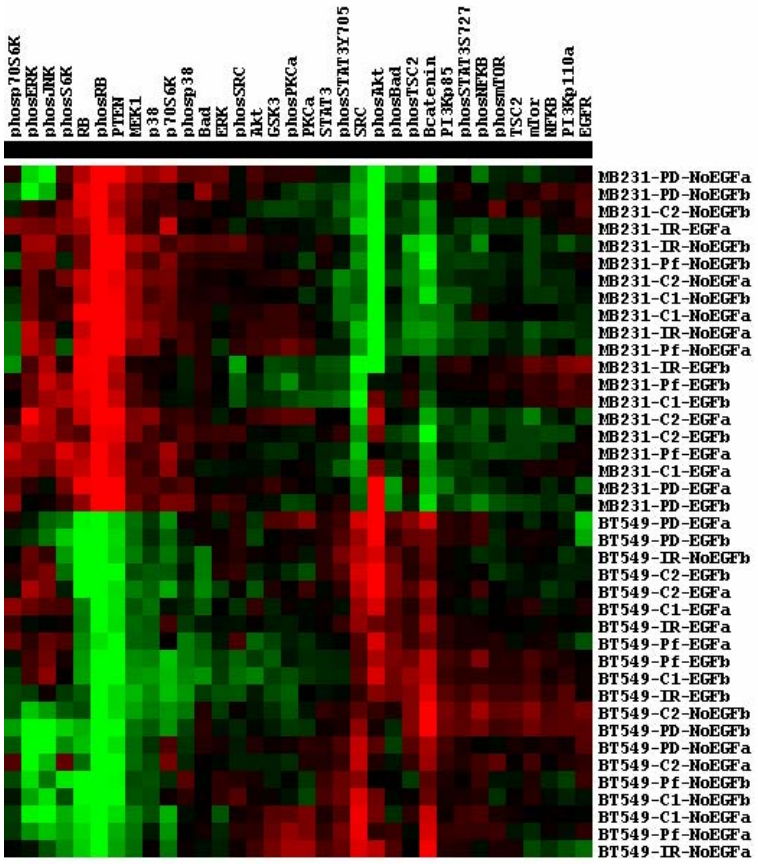


Figure 1. Reverse phase protein array data from BT549 and MDA-MB-231 breast tumor cells.

The cells were serum starved overnight and treated with the EGFR inhibitor Iressa, or the MEK inhibitor PD98059 or the AKT inhibitor Perifosine for 2 hours. Control cells were treated with DMSO. The cells were then stimulated with EGF or vehicle for 30 minutes and the cells were then lysed. The soluble proteins were spotted onto the reverse phase protein arrays, printed on Fast20

nitrocellulose coated slides and probed with antibodies to the different phosphor and total proteins of the EGFR network. The data shows the quantitative changes and is visualized in the heat map, where red indicates high and green indicates low. The lysates were probed with 40 different antibodies.

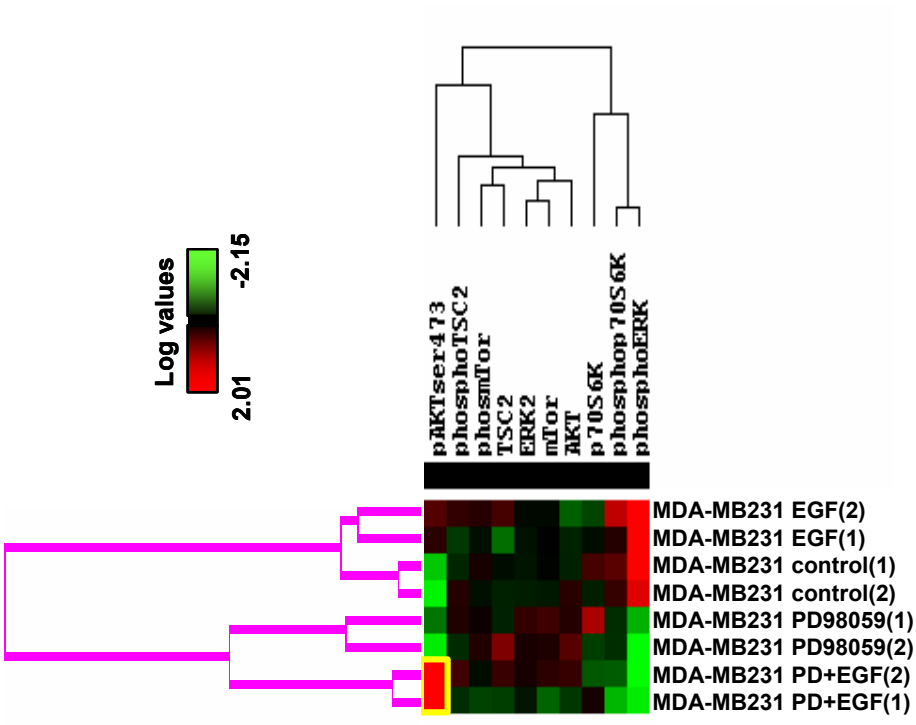


Figure 2.
Magnification of the MEK inhibitor data shows an increase in AKT phosphorylation when cells are treated with the MEK inhibitor.

Form the data we observed that MEK inhibition was leading to an increase in AKT phosphorylation. We tested this in a panel of breast tumor cell lines and we observed that in all cell lines inhibiting MEK increased AKT phosphorylation.

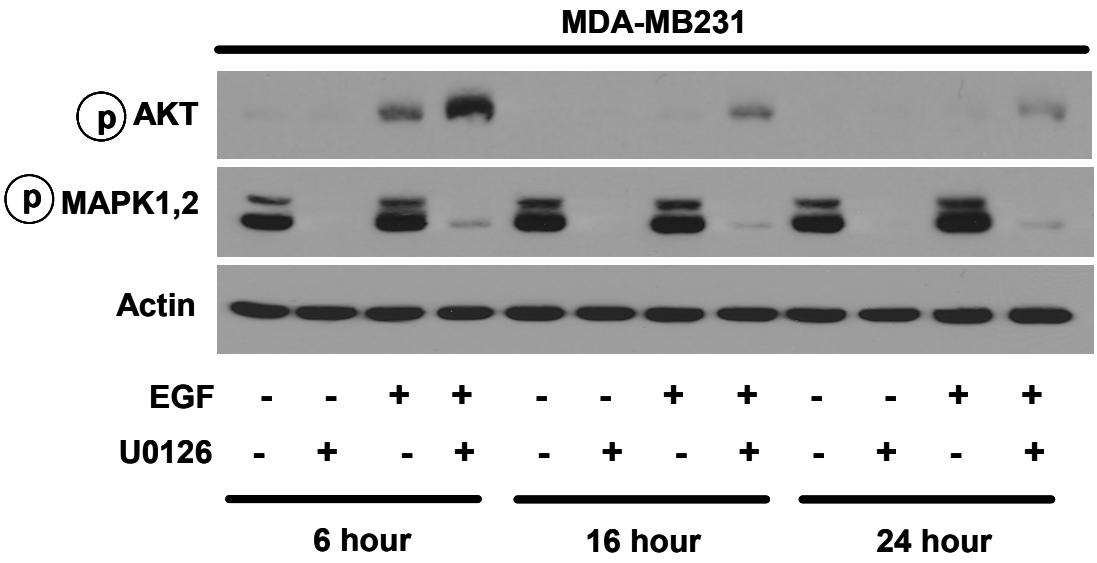


Figure 3. Time course of the MEK inhibitor shows that the increase in AKT phosphorylation is present upto 24 hours

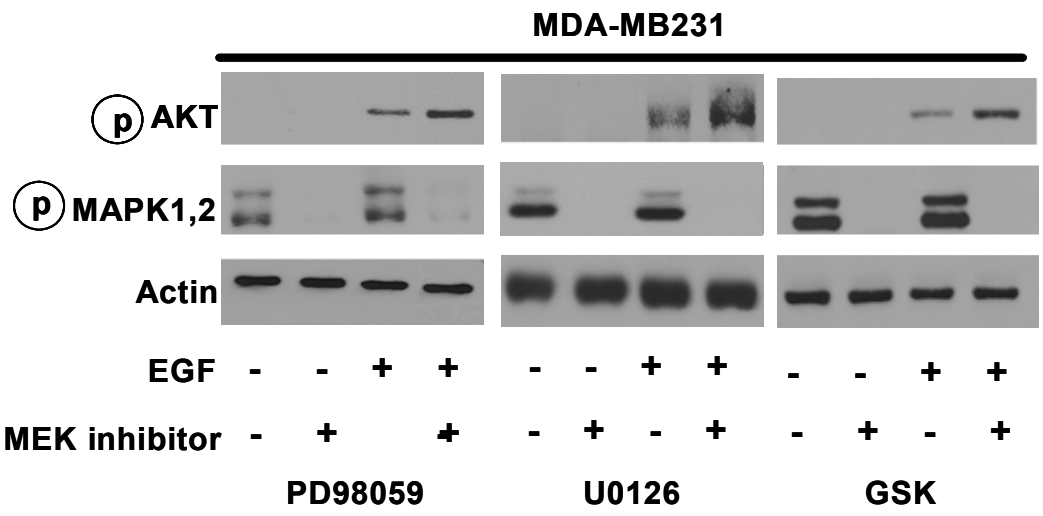


Figure 4. The increase in AKT is present with three structurally different MEK inhibitors

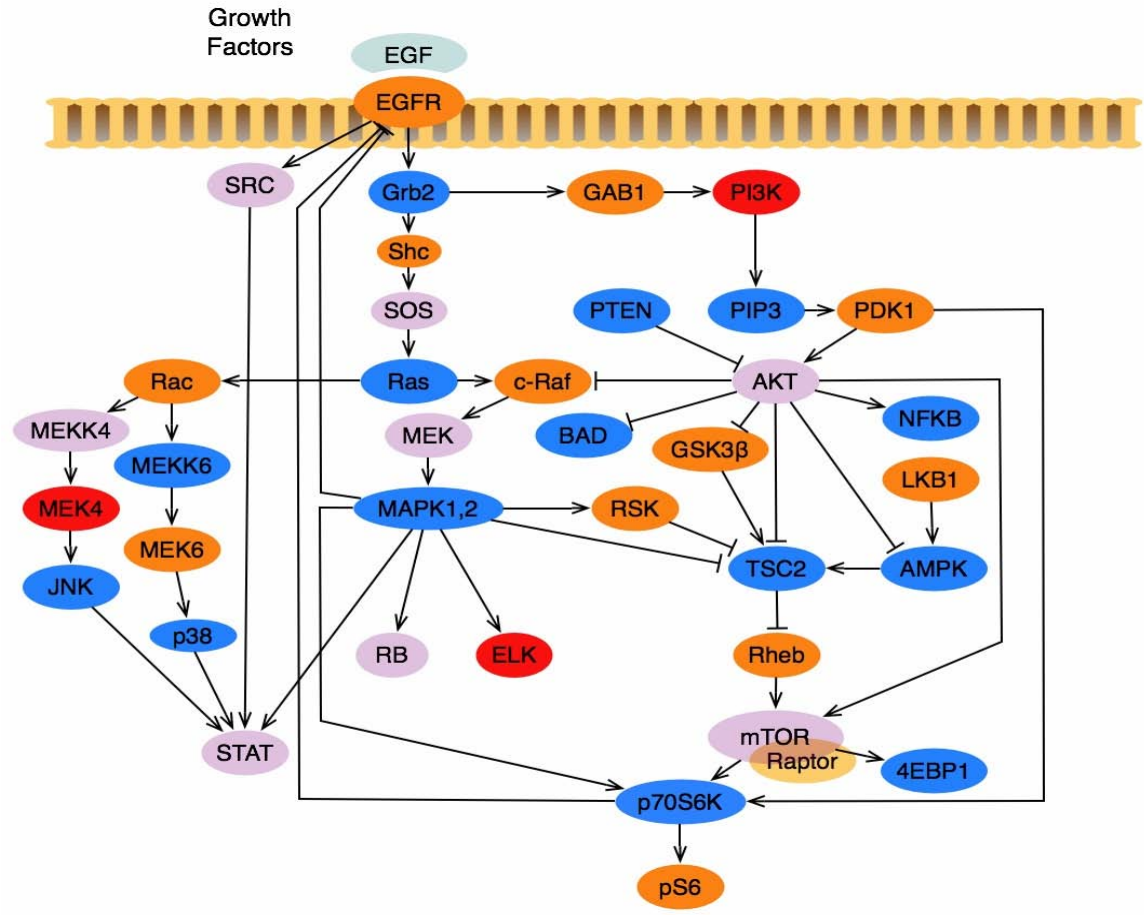


Figure 5. We developed a computational model of a EGFR-MEK-MAPK-AKT network and simulated regulatory loops that could increase AKT phosphorylation when MEK was inhibited.

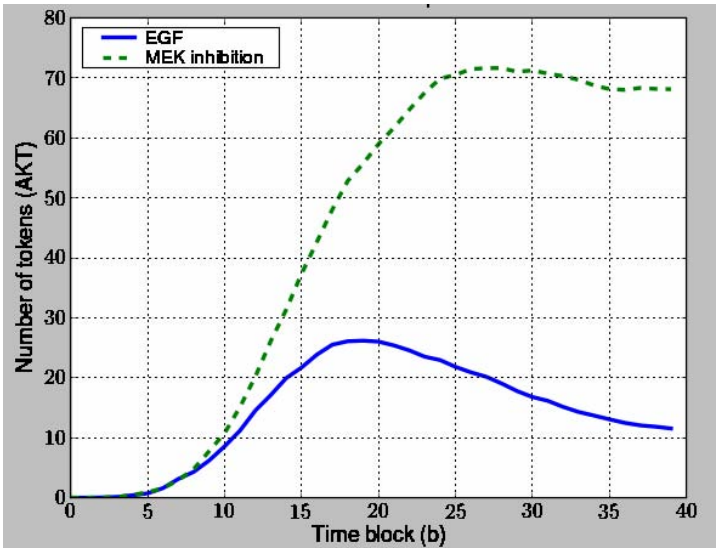


Figure 6. Computational model prediction of changes in AKT phosphorylation in the presence of MEK inhibitor.
Using the computational model we predicted the response of AKT to MEK inhibition and compared the model to experimental data. We observed that the model could predict the relative increases/decreases correctly but not the amplitude of the changes. To model the amplitude also will entail detailed biochemistry, instead we have decided for now to see if the model as it stands can have a functional predictive use.

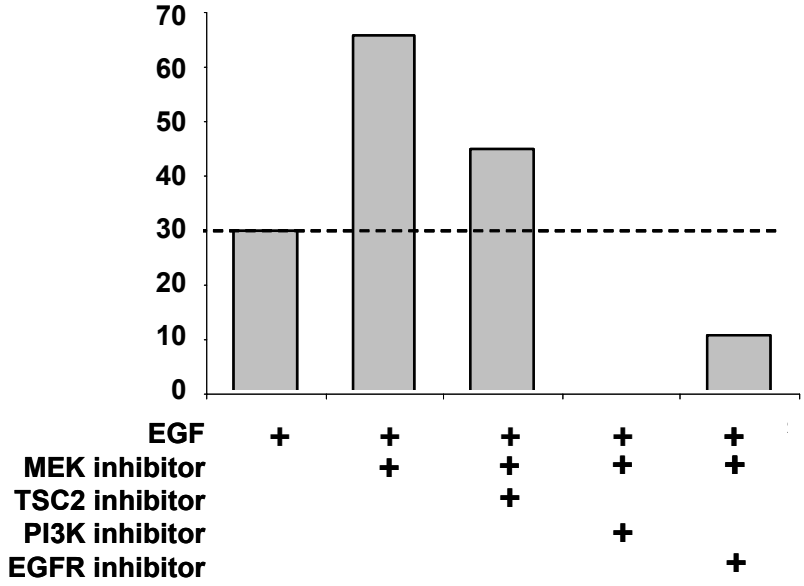


Figure 7. Modeling prediction of combinations to inhibit MAPK and AKT .
We used the model to predict combinations that can be used with the MEK inhibitor such that AKT is not increased. We tested these predictions on cell lines and assayed changes in proliferation.

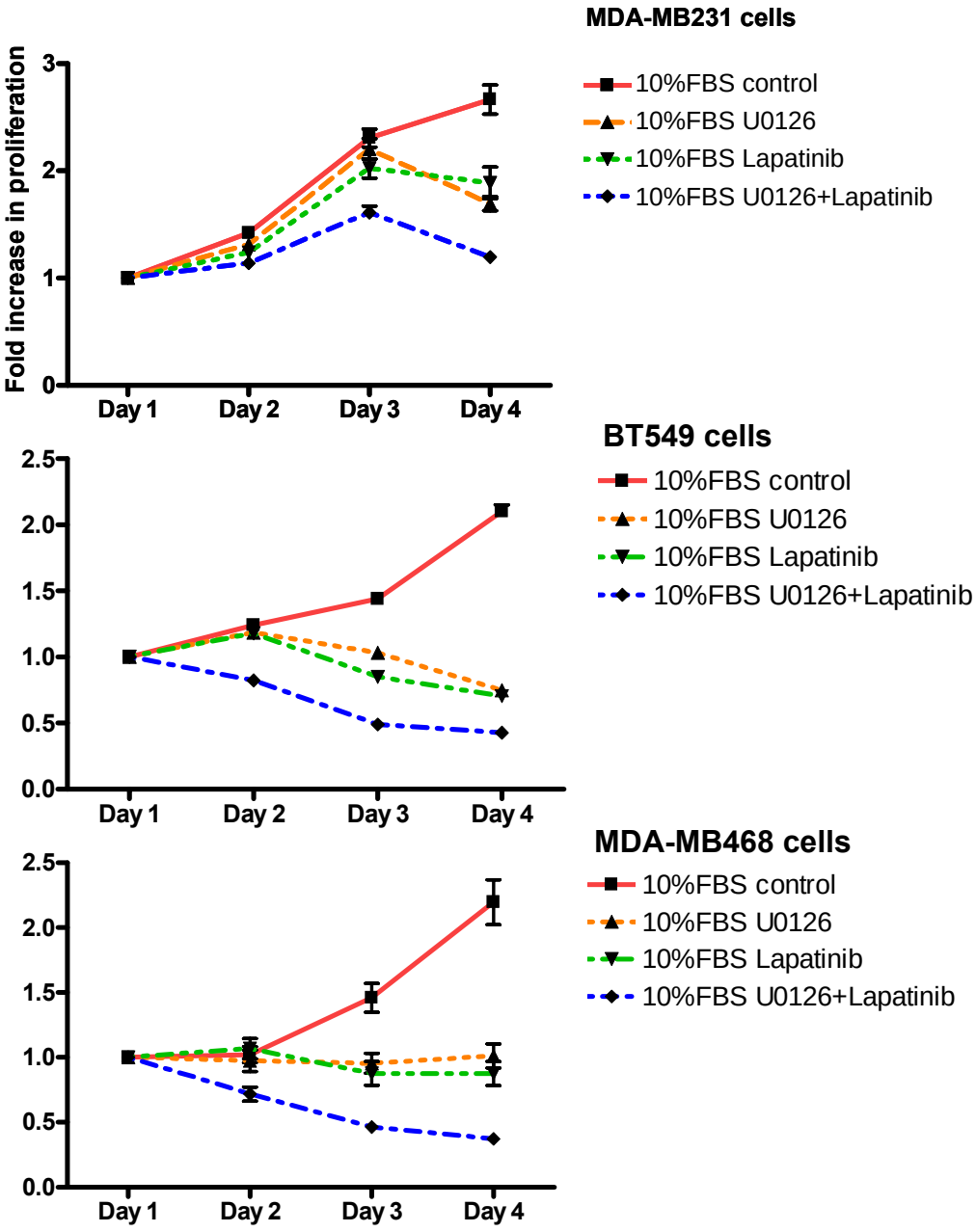


Figure 8. Experimental testing of modeling predictions.

We measured changes in proliferation with the different combinations and observed that a MEK inhibitor in combination with EGFR inhibitor decreased proliferation in cell lines with either Ras mutation or PTEN loss. Currently we are testing the other prediction from the model.

Key research accomplishments
Task 3. We have developed and made predictions from our model to identify combination of targets and are experimentally testing these predictions.
.
Reportable outcomes
From the work that we have done in the past year we have 3 papers published and 2 manuscripts submitted. In addition the support of this DOD grant was instrumental in obtaining a R01 grant from the NCI on targeted therapy in breast cancer.

Conclusions
We have developed a computational model of the signaling network. We have also generated data of the dynamic changes in signaling and integrated the biological data with the computational model. We have expanded the initial plan to include pharmacological inhibitors in addition to siRNA and are currently finishing up the final aim of the project.

Reference (papers from our work published/submited this year)

Ruths D, Muller M, Tseng J-T, Nakhleh L, Ram PT The signaling Petri Net-based simulator: A non-parametric strategy for characterizing the dynamics of cell-specific signaling networks PLoS Comp Biol 2008 Feb 29;4(2):e1000005

Muller M, Obeyesekere M, Mills GM, Ram PT Network topology determines dynamics of the mammalian MAPK1,2 signaling network: bi-fan motif regulation of C-Raf and B-Raf isoforms by FGFR and MC1R. FASEB J 2008 22:1393-403

Ruths D, Nakhleh L, Ram PT Rapidly Exploring Structural and Dynamic Properties of Signaling Networks Using PathwayOracle BMC Systems Biology 2008 In Press

Melissa Muller, Derek Ruths, Jen-Te Tseng, Richard Wooster, Barbara Weber, Sylvie Lequerre, Jeffrey Jackson, Luay Nakleh, Gordon B. Mills, Prahlad T. Ram MEK inhibitors disrupt homeostatic regulatory loops through EGFR to increases AKT phosphorylation in breast cancer cells. Submitted to Cancer Research

Appendix
Ruths D, Muller M, Tseng J-T, Nakhleh L, Ram PT The signaling Petri Net-based simulator: A non-parametric strategy for characterizing the dynamics of cell-specific signaling networks PLoS Comp Biol 2008 Feb 29;4(2):e1000005

Muller M, Obeyesekere M, Mills GM, Ram PT Network topology determines dynamics of the mammalian MAPK1,2 signaling network: bi-fan motif regulation of C-Raf and B-Raf isoforms by FGFR and MC1R. FASEB J 2008 22:1393-403

Ruths D, Nakhleh L, Ram PT Rapidly Exploring Structural and Dynamic Properties of Signaling Networks Using PathwayOracle BMC Systems Biology 2008 In Press

OPEN ACCESS Freely available online

PLoS COMPUTATIONAL BIOLOGY

The Signaling Petri Net-Based Simulator: A Non-Parametric Strategy for Characterizing the Dynamics of Cell-Specific Signaling Networks

Derek Ruths[1]*, Melissa Muller[2], Jen-Te Tseng[2], Luay Nakhleh[1], Prahlad T. Ram[2]

1 Department of Computer Science, Rice University, Houston, Texas, United States of America, **2** Department of Systems Biology, University of Texas M. D. Anderson Cancer Center, Houston, Texas, United States of America

Abstract

Reconstructing cellular signaling networks and understanding how they work are major endeavors in cell biology. The scale and complexity of these networks, however, render their analysis using experimental biology approaches alone very challenging. As a result, computational methods have been developed and combined with experimental biology approaches, producing powerful tools for the analysis of these networks. These computational methods mostly fall on either end of a spectrum of model parameterization. On one end is a class of structural network analysis methods; these typically use the network connectivity alone to generate hypotheses about global properties. On the other end is a class of dynamic network analysis methods; these use, in addition to the connectivity, kinetic parameters of the biochemical reactions to predict the network's dynamic behavior. These predictions provide detailed insights into the properties that determine aspects of the network's structure and behavior. However, the difficulty of obtaining numerical values of kinetic parameters is widely recognized to limit the applicability of this latter class of methods. Several researchers have observed that the connectivity of a network alone can provide significant insights into its dynamics. Motivated by this fundamental observation, we present the signaling Petri net, a non-parametric model of cellular signaling networks, and the signaling Petri net-based simulator, a Petri net execution strategy for characterizing the dynamics of signal flow through a signaling network using token distribution and sampling. The result is a very fast method, which can analyze large-scale networks, and provide insights into the trends of molecules' activity-levels in response to an external stimulus, based solely on the network's connectivity. We have implemented the signaling Petri net-based simulator in the PathwayOracle toolkit, which is publicly available at http://bioinfo.cs.rice.edu/pathwayoracle. Using this method, we studied a MAPK1,2 and AKT signaling network downstream from EGFR in two breast tumor cell lines. We analyzed, both experimentally and computationally, the activity level of several molecules in response to a targeted manipulation of TSC2 and mTOR-Raptor. The results from our method agreed with experimental results in greater than 90% of the cases considered, and in those where they did not agree, our approach provided valuable insights into discrepancies between known network connectivities and experimental observations.

Citation: Ruths D, Muller M, Tseng J-T, Nakhleh L, Ram PT (2008) The Signaling Petri Net-Based Simulator: A Non-Parametric Strategy for Characterizing the Dynamics of Cell-Specific Signaling Networks. PLoS Comput Biol 4(2): e1000005. doi:10.1371/journal.pcbi.1000005

Editor: Satoru Miyano, The University of Tokyo, Japan

Received September 18, 2007; **Accepted** January 18, 2008; **Published** February 29, 2008

Funding: DR and LN were supported in part by a Seed Grant awarded to LN from the Gulf Coast Center for Computational Cancer Research, funded by John and Ann Doerr Fund for Computational Biomedicine. J-TT was supported in part by a training fellowship from the Pharmacoinformatics Training Program of the Keck Center of the Gulf Coast Consortia (NIH grant 5 T90 DK070109-03), and MM and PTR were supported in part by Department of Defense grant BC044268 to PTR.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: druths@rice.edu

Introduction

Signaling networks are complex, interdependent cascades of signals that process extracellular stimuli, received at the plasma membrane of a cell, and funnel them to the nucleus, where they enter the gene regulatory system. These signaling networks underlie how cells communicate with one another, and how they make decisions about their phenotypic changes, such as division, differentiation, and death. Further, malfunction of these networks may alter phenotypic changes that cells are supposed to undergo under normal conditions, and potentially lead to devastating consequences on the organism. For example, altered cellular signaling networks can give rise to the oncogenic properties of cancer cells [1,2], increase a person's susceptibility to heart disease [3], and have been shown to be responsible for many other devastating diseases such as congenital abnormalities, metabolic disorders and immunological abnormalities [1,4].

In light of the crucial role signaling networks play in the proper functioning of cells and biological systems as a whole, and given the grave consequences their alterations may have on the behavior of cells, elucidating the connections in the networks, and understanding how they operate, are two central questions in cell biology. However, unlike the "pathway view" of signaling as linear cascades, signaling networks are highly interconnected, involve cross-talk among several pathways, and contain feedback and feed-forward loops [5]. Figure 1 illustrates this issue in a network of signaling cascades, which is stimulated by EGF and contains several players in cancer pathways. For example, multiple paths lead from EGFR to mTOR-Raptor, resulting in feed-forward loops. Some of these paths activate mTOR-Raptor,

Author Summary

Many cellular behaviors including growth, differentiation, and movement are influenced by external stimuli. Such external stimuli are obtained, processed, and carried to the nucleus by the signaling network—a dense network of cellular biochemical reactions. Beyond being interesting for their role in directing cellular behavior, deleterious changes in a cell's signaling network can alter a cell's responses to external stimuli, giving rise to devastating diseases such as cancer. As a result, building accurate mathematical and computational models of cellular signaling networks is a major endeavor in biology. The scale and complexity of these networks render them difficult to analyze by experimental techniques alone, which has led to the development of computational analysis methods. In this paper, we present a novel computational simulation technique that can provide qualitatively accurate predictions of the behavior of a cellular signaling network without requiring detailed knowledge of the signaling network's parameters. Our approach makes use of recent discoveries that network structure alone can determine many aspects of a network's dynamics. When compared against experimental results, our method correctly predicted 90% of the cases considered. In those where it did not agree, our approach provided valuable insights into discrepancies between known network structure and experimental observations.

while others inhibit it. Further, the network contains two feedback loops, one from p70S6K to EGFR and another from MAPK1,2 to EGFR.

These and other complexities make it very difficult to analyze signaling networks by experimental biology approaches alone. As a result, computational methods have been developed and combined with experimental biology approaches, producing powerful tools for the analysis of these networks [6]. These computational methods produce hypotheses that guide the experimental design, leading to more informative experiments, while experimental results help refine the computational models, resulting in more accurate predictive tools.

In a recent survey, Papin et al. classified existing computational methods into two categories: *structural* and *dynamic* network analysis [6]. Structural network analysis is mainly based on the network's connectivity, which is typically readily available from numerous public signaling network databases (e.g., [7–9]), and makes inferences about global network properties as well as individual protein functions. This category can be further refined into two sub-categories, both of which are solely based on connectivity information, yet differ in the type of answers they provide. For example, the methods described in [10–13] infer "static" properties of the network, such as numbers of paths, reachability results, etc. In a series of papers, Palsson and co-workers [6,14–16] introduced extreme pathway analysis techniques, which are more appropriate for metabolic networks, yet have been applied to signaling networks to characterize various properties of networks, such as redundancy and cross-talk. Similar analyses have also been

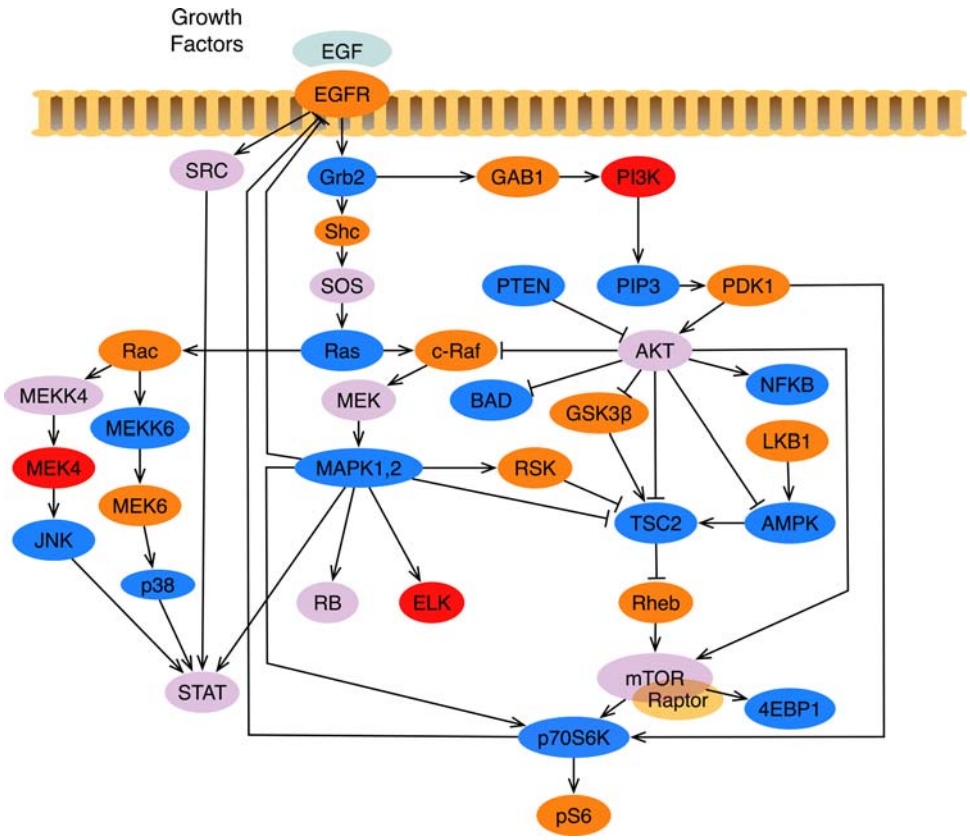


Figure 1. The Model Signaling Network. A MAPK1,2 and AKT network downstream from EGFR, which we assembled from various sources, and used for the case study analysis in this work. An edge from u to v ending with an arrow indicates an activating reaction, while an edge ending with a plunger indicates an inhibiting reaction. With the exception of TSC2, all nodes have self-inhibitory edges, which were added to model the external cellular machinery that regulates the concentration of the active form of the proteins [36–43]. Colors were selected to enhance readability of the network.
doi:10.1371/journal.pcbi.1000005.g001

formalized and conducted using the principles of S- and T-invariants in Petri Nets (e.g., [17–20]).

Methods for dynamic network analysis use, in addition to the network connectivity, the kinetic parameters of the biochemical reactions. The goal of these methods is to model the actual kinetics of the network and obtain through simulation the actual quantities of proteins involved in signal transduction. One of the most widely used techniques in this category is systems of ordinary differential equations (ODEs) (e.g., [21–25]). Within such a system, each reaction is modeled by a series of equations connecting reactant concentrations to product concentrations through differential relationships involving reaction rate constants. Given the difficulty of obtaining the numerical values of kinetic parameters [19,26] and standardization of the parameters and models [27], the applicability of these methods is limited in practice to small-scale networks [6,28].

Petri Nets have also been used for simulating the dynamics of signaling networks [29–31]. While such approaches somewhat relax the necessity for biologically exact kinetic parameters, current Petri Net-based approaches still require the selection of weights and/or probability distributions for individual interactions in the model. As a result, selecting the values for Petri Net parameters presents challenges similar to those encountered in ODE modeling.

Structural network analysis assumes mainly connectivity information about the model, and provides insights into global, static properties of the network. Dynamic analysis in general assumes numerical values of the kinetic parameters, and provides predictions of network dynamics by quantifying the change in concentration and activity-level (the concentration of the active form of a given protein) of the individual proteins and complexes in the network. To obtain a more detailed analysis one must either solve parameter optimization problems for a large number of molecules and interactions or conversely experimentally derive these values.

Given the difficulty of obtaining numerical values of kinetic parameters [19,26] and the implications this has on the applicability of dynamic analysis methods [6], it is imperative to develop innovative approaches that combine the attractive low requirements of structural network analysis techniques with the detailed answers provided by dynamic analysis techniques—specifically the response of individual proteins to signals which travel through the network.

Several recent efforts in this direction have produced encouraging results. An approach using a boolean network simulation method, based on work in the area of gene regulatory networks, successfully used only signaling network connectivity information to predict the speed of signal transduction through a stomata signaling network [32]. The use of piecewise linear systems of ODEs have also had success in analyzing some of the dynamics of gene regulatory and signaling networks without using exact kinetic parameters (e.g., [33–35]). The obstacle to extending the method in [32] to model individual protein responses to signal transduction is the boolean model used to discretize the signal as it propagates. In a boolean model, the signal is either present or absent at each node in the network. Such two-state models of signal transduction simplify the underlying biochemistry to the point where it is difficult to model changes in protein concentration more precisely than present or absent. Modeling such gradients of concentration changes and the effects of those changes may be important to predicting individual protein responses, motivating our effort to devise more fine-grained ways to model and simulate the dynamics of signaling networks. The challenges to using linear-piecewise ODEs to model a signaling network center around the issue of identifying all the ODEs required to model the underlying network as well as scalability issues involved in simulating large systems of ODEs.

In this paper, we extend the synchronized Petri net model and firing policy such that the resulting framework models cellular signaling processes. We call this extension the signaling Petri net (SPN). By coupling this with a novel strategy for Petri net execution and sampling, we obtain a method capable of characterizing some dynamics of signaling networks while using only connectivity information about these networks.

To validate our method, we studied the MAPK1,2 and AKT network shown in Figure 1 in two breast cancer cell lines. This network was chosen because the EGFR receptor and its downstream signaling network play a very important role in development, differentiation, and oncogenic transformation. Two very important signaling molecules within the cell are MAPK and AKT, both of which can be activated by EGFR, and contains several potential regulatory paths between them. We constructed a model network of EGF regulation of MAPK and AKT which includes several feedback and feed-forward loops all of which were constructed based on experimental findings from different laboratories around the world [36–43]. We analyzed, both experimentally and computationally, the change in activity-level of several proteins in response to targeted manipulation of TSC2 and mTOR-Raptor. Using the model network, the predictions from our method agreed with experimental results in over 90% of the cases, and in those where they did not agree, our method correctly identified discrepancies that could be traced back to incompleteness in the network connectivity model.

Materials and Methods

Our approach combines elements of the boolean network simulator in [18] with a synchronized Petri net model [44]. In [18], Li et al. present a non-parametric approach that accurately predicts the speed of signal propagation through a network. However, as their method assumes a binary model of activation—every protein is either active (*true*) or inactive (*false*)—modeling a range of activity-levels is difficult. Petri nets, while able to model concentrations using tokens, require parameters describing the kinetic characteristics of the network, which are typically difficult to obtain.

Our method models signal flow as the pattern of token accumulation and dissipation within places (proteins) over time in the Petri net. Transitions in the network represent directed protein interactions; each transition models the effect of a source protein on a target protein. Through transition firings, the source can influence the number of tokens assigned to the target, called the *token-count*, modeling the way that signals propagate through protein interactions in cellular signaling networks.

In order to overcome the issue of modeling reaction rates in the network, signaling dynamics are simulated by executing the signaling Petri net (SPN) for a set number of steps (called a *run*) multiple times, each time beginning at the same initial marking. For each run, the individual signaling rates are simulated via generation of random orders of transition firings (interaction occurrences). When the results of a large enough number of runs are averaged together, we find that the series of token-counts correlate with experimentally measured changes in the activity-levels of individual proteins in the underlying signaling network. In essence, the tokenized activity-levels computed by our method should be taken as abstract quantities whose changes over time correlate to changes that occur in the amounts of active proteins present in the cell. It is worth noting that some of the most widely used experimental techniques for protein quantification—western blots and microarrays—also yield results that are treated as indications, but not exact measurements, of protein activity-levels within the cell. Thus in some respects, the predictions returned by

our SPN-based simulator can be interpreted like the results of a western blot or microarray experiment looking at changes relative to "control".

The key insight behind our approach is the assumption that, while all network parameters determine the actual signal propagation to some extent, the network connectivity is the most significant single determinant. While this is clearly a gross simplification, several researchers have observed that the connectivity of a biological network dictates, to a great extent, the network's dynamics [18,45–47]. Some have conjectured that biological network connectivities have evolved to have a stabilizing effect on the overall network behavior, making the network more resilient to local fluctuations in other network parameters such as reaction rates and protein binding affinities [45,47]. Here we present the *signaling Petri net* (SPN) model and the signaling Petri net-based simulator whose designs collectively utilize this assumption and couple it with a Petri net tokenization scheme that quantifies the changes in protein activity-levels that occur as signals propagate through the network. In the following sections, we describe the synchronized Petri net, how we extended it to create the signaling Petri net, and a novel strategy for executing the signaling Petri net to simulate signaling network dynamics.

Petri Nets

A Petri net is a graph that consists of two types of nodes, *places*, and *transitions* [44]. Edges in the graph, called *arcs*, are directed and connect places to transitions or transitions to places. Thus, the Petri net is a bipartite graph. Formally, a Petri net is a 4-tuple $Q=\langle P,T,I,O\rangle$ where

$P=\{p_1,p_2,\ldots,p_m\}$ is the set of places,

$T=\{t_1,t_2,\ldots,t_n\}$ is the set of transitions,

$I=\{i_1,i_2,\ldots,i_k\}$ is the set of input arcs where for all $(u,v)\in I$, $u\in P$ and $v\in T$, and

$O=\{o_1,o_2,\ldots,o_l\}$ is the set of output arcs where for all $(u,v)\in I$, $u\in T$ and $v\in P$.

In order to simulate a dynamic process, a number of tokens is assigned to each place in order to indicate the presence of some quantitative property. This assignment of tokens to places encodes the state of the system and is called a marking, denoted $\mathbf{m}$. A *marked Petri net*, $R=\langle Q,\mathbf{m}_0\rangle$, is a Petri net with a marking $\mathbf{m}_0$, called the initial marking. For the remainder of this paper, the term *Petri net* (PN) refers to a marked Petri net.

Changes in the state of the system are simulated by *executing* the Petri net—evaluating the effect of transitions on the marking of the network. These changes in marking are induced by sequential *firing* one or more transitions. When a transition fires, it removes a token from each place connected to it by input arcs and adds a token to each place connected to it by output arcs. The number of tokens removed from inputs and added to outputs can be specified by weighting the input arcs. However, as our extension does not use this weighting property, we do not consider this very common PN formulation here.

A transition can only fire when it is *enabled*, meaning that each of its input places has at least one token in the current marking. If a transition t, when fired on a marking $\mathbf{m}_1$, produces marking $\mathbf{m}_2$, then we write $\mathbf{m}_1|t\rangle\mathbf{m}_2$.

This notation can be extended to represent the effect of firing a series of transitions. A *firing sequence*, $\sigma=(t_1,t_2,\ldots,t_j)$ is a sequence of transitions. The sequence's cumulative effect on the system's state is denoted $\mathbf{m}_0|\sigma\rangle\mathbf{m}_f$ where $\mathbf{m}_0$ is the initial marking and $\mathbf{m}_f$ is the marking produced by the firing of the sequence of transitions in the order specified in σ. In this paper, we write $\boldsymbol{m}_g^{\sigma}$ to indicate the marking produced by the first g transitions in σ. Therefore, in the above example, $\boldsymbol{m}_0^{\sigma}=\boldsymbol{m}_0$ and $\boldsymbol{m}_{|\sigma|}^{\sigma}=\boldsymbol{m}_f$.

For a more complete introduction to types of Petri nets and their properties, we refer the reader to [44].

Synchronized Petri nets. Synchronized Petri nets model systems in which the firing of a transition is triggered by a specific event that occurs in the environment. The marked Petri net is extended to include a set of these events and a mapping function that assigns an event to each transition. When transition t's assigned event occurs, transition t is fired. Formally, a synchronized Petri net is a 3-tuple $\langle R,E,Sync\rangle$, where [44]:

R is a marked Petri net,

$E=\{e_1,e_2,\ldots,e_s\}$ is a set of events, and

$Sync: T\rightarrow E\cup\{\mathbf{e}\}$ maps each transition in the Petri net to an event. Event $\mathbf{e}$ is the *always occurring event*. Any transition associated with $\mathbf{e}$ is always immediately fired upon becoming enabled.

When executing a synchronized Petri net, transition t is fired when its associated event $e=Sync(t)$ occurs. The order in which events are generated depends upon the environment which generates them. Just as in the marked Petri net, when a transition fires, it removes one token from each place connected by input arcs and gives one token to each place connected by output arcs.

As will be discussed in the next sections, we extend the synchronized Petri net paradigm to model the dynamics of a signaling network. To our knowledge, ours is the first use of the synchronized Petri net to model biochemical systems. In principle it is well suited to signaling networks since places represent proteins, tokens represent concentrations, and transitions represent directed protein interactions. A model of signaling event occurrence can be used to generate events and fire transitions, providing a way of simulating the signaling network's behavior. These and other design details will be discussed in the next section.

The Signaling Petri Net-Based Simulator

A high-level sketch of our simulator is given is Figure 2. Details and rationale for specific design decisions will be discussed in subsequent sections.

During the simulation, the input signaling Petri net is executed multiple times on a firing sequence constructed by the signaling event generator. The signaling event generator imposes an ordering on transition firing such that it creates a two-time scale simulation. The smaller time scale is discretized as the firing of a single transition. This unit is referred to as the *firing* time scale. Firing steps are nested within a larger time scale, called time *blocks*, in which each transition is fired exactly once. Thus, there are $|T|$ firings per block. Since the simulation is run for the specified number of time blocks, B, there are $B|T|$ firing steps in the simulation.

The time structure for an example simulation is illustrated in Figure 3. This dual-time approach is necessitated by the rate parameter sampling strategy we employ. Since the rate parameters are not known, our method executes many simulation runs (Step 2 in Figure 2) in order to sample the space of possible rate parameters. The markings returned by these runs are then averaged (Step 3 in Figure 2). The only requirement placed on the different rate parameter values is that all events occur within the same larger time frame—the time block. Therefore, within every time block all edges are evaluated once, though not necessarily in the same order.

This idea of evaluating random event orderings within a two-time scale system has appeared before in the domain of transcriptional networks [48]. In that study, Chaves et al. employed a two-time scale formulation of network updates similar in concept to the one we describe here. In their work, they assumed a boolean model of regulation and characterized the

PROCEDURE SIMULATE(S, B, r)

1. Set the initial marking of S
2. For $i = 1$ to r
   (a) Generate a random sequence of the events in E
   (b) Simulate the network by executing the transitions associated with the events in the generated sequence
   (c) Record the number of tokens at each node, for each time step
3. For each node, compute the number of tokens at each time unit t, averaged over r

Figure 2. A High-Level Outline of the Procedure for Simulating a Signaling Network. The input to the procedure is a signaling Petri net, S, the number of time units to simulate the network for, B, and the number of runs for which to repeat the simulation, r. The random generation of event ordering is employed to simulate the stochasticity in reaction rates and the differing times of signal arrivals.
doi:10.1371/journal.pcbi.1000005.g002

effect of different relative rates of transcription within the same network on the final steady state reached. In contrast, our method is designed to operate on tokenized models of signaling networks with the ultimate intent of predicting the activity-level changes of proteins in the underlying signaling network over time.

In the next sections, we discuss in greater detail the core design decisions underlying our method: the signaling Petri net, transition firing, signaling network event generator, constructing the initial marking for the model, and sampling signaling rates. We then discuss how our strategy can be used to predict the outcome of perturbation experiments.

The Signaling Petri Net

The goal of our method is to predict the signal flow through a cell-specific network under specific experimental conditions. As a result, the signaling Petri net model must characterize the connectivity of the signaling network, the connectivity-level network properties that are unique to the cell type and experimental conditions under which the network is being studied, and the signaling processes of activation and inhibition.

The signaling Petri net is a synchronized Petri net with: 1) a specific way of modeling activating and inhibiting interactions using places, transitions, and arcs; 2) a one-to-one correspondence between events and transitions such that every transition is associated with a unique event; 3) modified rules regarding how many tokens are moved in response to a transition firing; and 4) a signaling network event generator.

Places correspond to the activated forms of signaling proteins. The number of tokens assigned to place p in marking $\mathbf{m}_s$, $\mathbf{m}_s(p)$, abstractly represents the amount of active protein p present in that network state. Signaling interactions are modeled using transitions and their connected input and output arcs. Each transition, t, is associated with a unique signaling event, e, such that when e occurs, transition t fires. Figure 4 shows the equivalent signaling Petri net for a signaling network.

Formally, a signaling Petri net is a 3-tuple $S = \langle R, E, Sync \rangle$, where:

R is a marked Petri net,

E is a set of signaling events such that $|E| = |T|$ and there is no *always occurring event*, and

$Sync: T \rightarrow E$ is a one-to-one mapping which assigns each transition a unique signaling event.

The initial marking of a signaling Petri net, $\mathbf{m}_0$, represents the state of rest from which the network is starting and being simulated. Proteins whose concentrations are known to be high can be given a large number of tokens, and those whose concentrations are known to be low can be assigned few or zero tokens. Attention to the initial marking is central to modeling cell-specific networks. In many cell lines, specific proteins are known to contain mutations that render them perpetually active or inactive [49]. Furthermore, experimental studies frequently involve the targeted manipulation of various proteins within the network. Both of these phenomena induce state changes in certain proteins at various time points that must be modeled. The way in which these are modeled will be discussed when the simulator design is explained.

Transition Firing

When a signaling interaction $A \rightarrow B$ (A *activates* B) or A⊣B (A *inhibits* B) occurs, it has the effect of changing the state of the system by modifying the activity-level of A and/or B. Thus, in the SPN used to model this network, the associated transition, t, will fire at

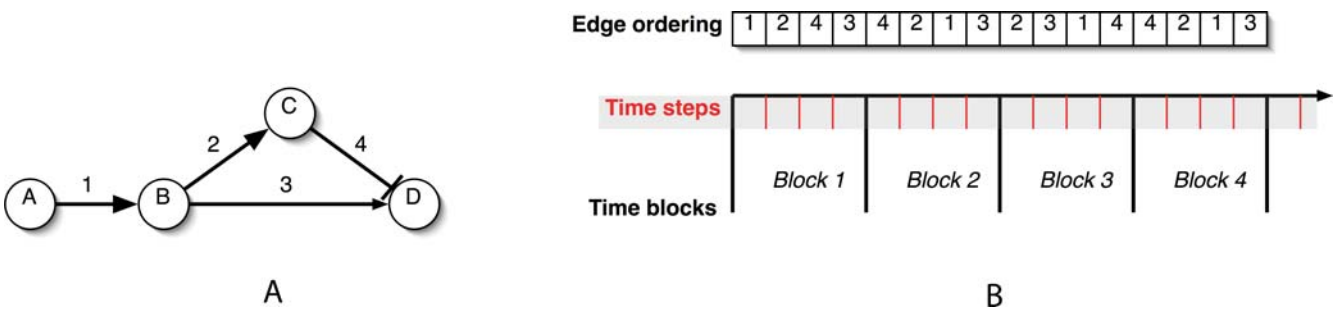


Figure 3. The Effects of Reaction Rates on Signal Propagation. (A) By changing the speed of signaling edge 3, the value of D at the end of a single simulation step can be reversed. If edge 3 is slower than the cascade B→C⊣D, then D will be active. If edge 3 is faster than the cascade, then D will be inactive. (B) An example of how the simulator might evaluate the individual edges during a run. In each time block, every edge is evaluated once. Each edge evaluation corresponds to one time step. Note that the order of the edge evaluation is shuffled during each time block in order to sample the space of possible relative signaling rates.
doi:10.1371/journal.pcbi.1000005.g003

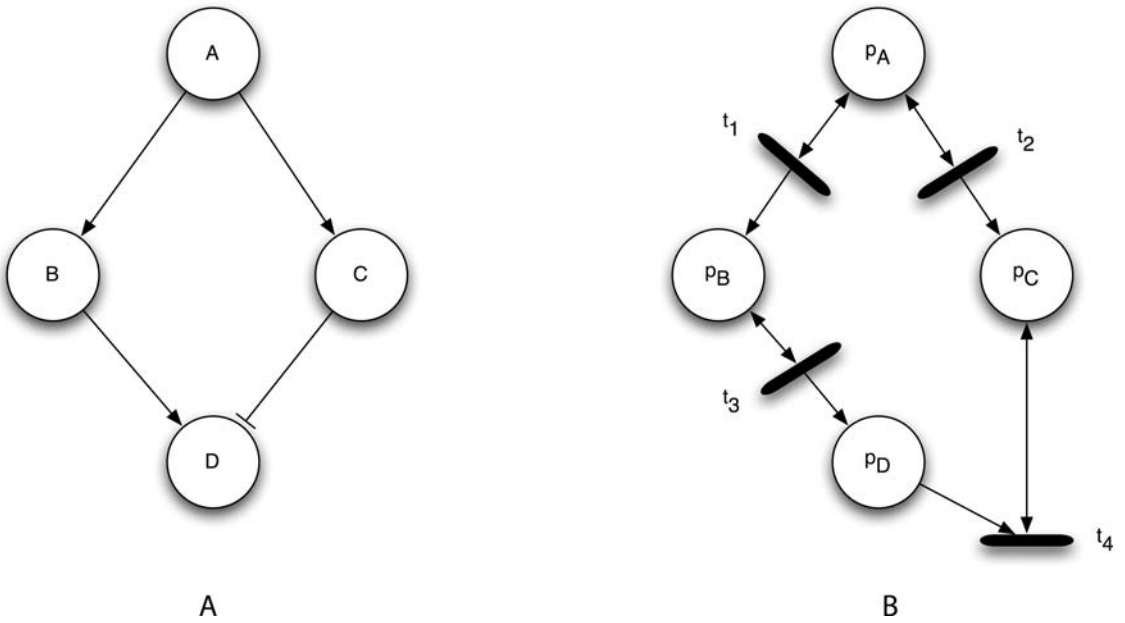


Figure 4. An Example Signaling Network and Its Corresponding Petri Net. An example signaling network (A) and its corresponding Petri net (B). Each signaling protein in the network, A, B, and C, are designated as places p_A, p_B, and p_C. Signaling interactions become a transition node and its input and output arcs. Note that the connectivity for an activating edge differs from that of an inhibitory edge.
doi:10.1371/journal.pcbi.1000005.g004

time τ and produce marking $\mathbf{m}_{\tau+1}$ from $\mathbf{m}_{\tau}$. The way in which $\mathbf{m}_{\tau+1}$ is computed from $\mathbf{m}_{\tau}$ depends on the set of input and output arcs attached to the transition as well as the number of tokens moved by the transition.

The combination of input and output arcs connected to a transition is determined exclusively by the type of interaction and the transition firing model. However, different topologies, combinations of input and output arcs, are needed to model the different biochemical processes that mediate protein-protein interactions in a signaling network. Here we examine four of the most common biochemical processes, identify the corresponding topological motifs, and ultimately devise a modeling policy best suited for non-parametric simulation of signal flow.

In *post-translational modification* (PTM), a protein mediates the addition or removal of a phospho group at a specific phosphorylation site on another protein. In *GTP/ATP* binding, a protein triggers the exchange of GDP (ADP) from GTP (ATP) on another protein. In a *recruitment* process, a protein mediates the relocalization of another protein to a different part of the cell. Finally, in a *complexing* process, a protein binds to another protein to create a complex, which can then participate in other reactions. In the first two processes, the mediating protein usually acts as an enzyme that participates in the reaction but is not consumed by the reaction. In the latter two processes, the participating protein often becomes unavailable to other reactions, transiently while the protein recruitment is taking place and for longer durations when complexing occurs. To model these two cases, we identified the two different token-passing policies implemented by the different topological motifs depicted in Figure 5.

Token consumption. In this policy, uΠv consumes tokens in u in order to generate new tokens for v. In order to model this, p_u is connected to transition t_1 through an arc and p_v is connected to t_1 through an output arc. When t_1 fires, some number of tokens in p_u are moved into p_v. Similarly, u$\dashv$v consumes tokens in u in order to consume tokens in v. This is modeled by connecting p_u to t_2 with an input arc and p_v to t_2 with an input arc. When t_2 fires, some number of tokens are removed from both p_u and p_v. This policy models a recruitment or complexing event in which u binds to another molecule, thereby creating a molecule of type v. *A molecule of type u has been consumed in order to generate or deactivate a molecule of type v.*

Token conservation. In this policy, uΠv generates new tokens for v while conserving those in u. In order to model this, p_u is connected to transition t_3 through a read arc. Node p_v is connected to t_3 through an output arc. When t_3 fires, some number of tokens in p_u is read (but not removed) and copied into p_v. Similarly, u$\dashv$v consumes tokens in v while conserving those in u. This is modeled by connecting p_u to t_4 with a read arc and p_v to t_4 with an input arc. When t_4 fires, some number of tokens in p_u are read and removed from p_v. Enzymes will often behave in this way: inducing a change in a molecule (v) without themselves undergoing any change. *A molecule of u has induced a change in a different molecule of type v without itself changing state.*

Ideally, for each interaction in the network, the associated transition could be embedded in the topology corresponding to the interaction's underlying biochemical mechanism. However, connectivity-level knowledge of the network does not provide this information for each interaction. In the absence of these details, we use one token-passing policy for all interactions in the network. We implemented and tested both the consuming and conserving policies and found that token conservation provides significantly more accurate results when compared to experimentally derived data. This is not surprising, as post-translational modification and GTP/ATP binding events are responsible for many activation state changes in signaling networks [1,50–52]. It is worth noting that our approach does not restrict the net structure to token conserving topologies. Thus, it is possible to use the token consumption topologies where such processes are known to occur. However, as our focus in this paper is designing a purely non-parametric simulation method, we consider the use of information regarding the biological mechanism of signaling as a potential way to further improve the accuracy of our method's predictions and identify this as a direction for future work.

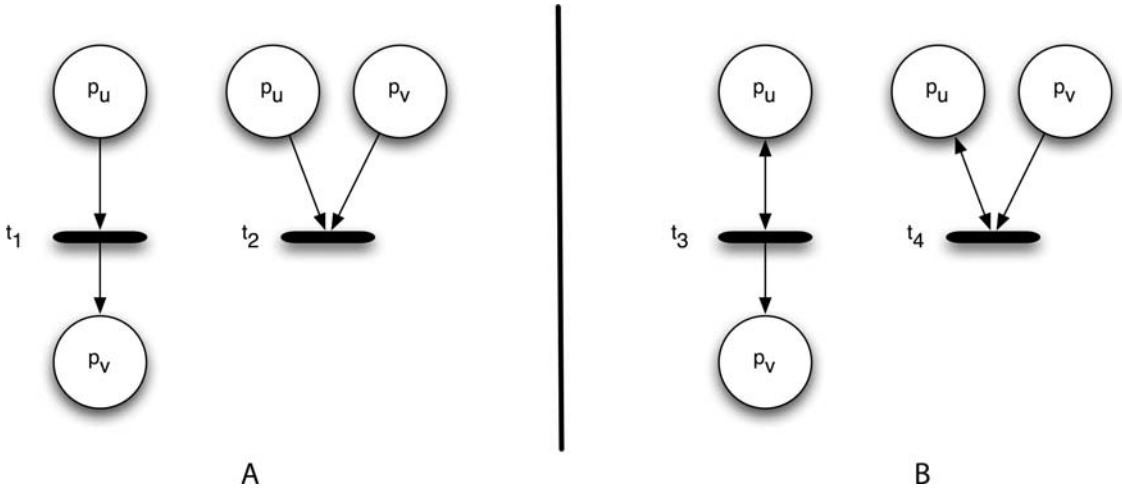


Figure 5. The Topological Motifs for Differing Signaling Processes. (A) The token consumption motifs for complexing and recruitment. Transition t_1 encodes activation of v by the binding or consumption of u. Transition t_2 encodes deactivation of v by the binding or consumption of u. In both cases, the number of tokens of p_u decreases immediately after transitions t_1 and t_2 fire. (B) The token conserving motifs for PTM and GTP/ATP binding. Transition t_3 encodes enzymatic activation of v by u. Transition t_4 encodes enzymatic inhibition of v by u. In both cases, the number of tokens of p_u remains unchanged immediately after transitions t_3 and t_4 fire.
doi:10.1371/journal.pcbi.1000005.g005

The transition topologies, as described above, do not designate how the number of tokens added to or removed from p_v is determined. However, we know that in biochemical signaling networks concentration has an effect on the strength of a signaling event [53–55]. Specifically, the higher u's concentration, the stronger its effect on v—the more tokens that p_u has, the more tokens of p_v should be affected (generated or consumed).

However, because of the stochastic nature of the underlying biochemistry, it would be inaccurate to assume that *all* active u molecules will always participate in an interaction with v. In order to accommodate this observation, when transition t fires, we randomly select the number of p_u's tokens to be involved in the subsequent evaluation of the transition, which we call a *signaling event*. Note that, according to our choice of topology, p_u can always be identified as the node connected to the transition by a read arc. In this paper, we assume a uniform distribution for selecting the number of tokens involved in a given signaling event, but acknowledge that other distributions may be more appropriate under certain circumstances and identify this as a topic deserving further consideration.

Let $m_s(x)$ denote the number of tokens in node x at time s. For an interaction (u,v), under the token conservation policy detailed above, u's token-count remains unchanged after the firing of t, whereas v's token-count is updated based on the following formula:

$$m_s(v) = \begin{cases} m_{s-1}(v) + random(0, m_{s-1}(u)) & \text{if } u \text{ activates } v \\ \max\{0, m_{s-1}(u) - random(0, m_{s-1}(u))\} & \text{if } u \text{ inhibits } v \end{cases},$$

where $random$(p,q) is a random integer drawn from a uniform distribution over the range [p,q].

If we employ the policy of token passing with consumption, then after $m_s(v)$ has been computed based on the formula above, $m_s(u)$ is updated as:

$$m_s(u) = m_{s-1}(u) - \min\{m_{s-1}(u), |m_s(v) - m_{s-1}(v)|\}.$$

Signaling Network Event Generator

The SPN topology and transition token-number selection policy alone do not specify the speed with which individual signaling interactions occur. However, such rates must be accounted for when simulating a signaling network. ODEs characteristically model such details as reaction rate constants; parameterized Petri nets specify these in a variety of ways including transition firing rates and firing probabilities [17,30]. In synchronized Petri nets, the environment controls the generation of events. Thus, the signaling network event generator is responsible for controlling the timing and ordering of signaling events. However, as our objective is a non-parametric simulation method, our approach must either estimate these parameters or operate without explicit knowledge of them.

Estimating reaction rates using only connectivity is currently beyond the predictive or inferential capabilities of computers. While there has been some work in the area of predicting reaction rates, all results of which we are aware require knowledge about the mechanism of signaling (e.g., [56]). As a result, without enriching the SPN model, it is doubtful that rate parameters can be accurately estimated.

For this reason, the signaling network event generator operates without explicit knowledge of the rate parameters. To compensate for this "missing" knowledge, we make use of an observation of signaling networks discussed earlier: a network's connectivity determines its dynamics. Several studies have found that the connectivity of biochemical networks desensitizes them to small fluctuations in the kinetic biochemical parameters [45–47]. Understood within the context of evolution – a stochastic process that tweaks signaling network parameters across generations – this is a highly desirable property as it ensures that an offspring remains viable despite fluctuations in the exact tuning of its cellular machinery. If this property holds, then small fluctuations in the rate parameters should have a marginal effect on the overall propagation of signal through the network. We can consider these small effects to be noise obscuring the underlying dynamics of the network connectivity. By taking many samples of the network dynamics under a variety of reaction rate assignments and then averaging these dynamics, we simultaneously reduce the noise

introduced by any one rate assignment and strengthen the underlying dynamic characteristics of the network's connectivity.

However, since reaction rate constants can vary by several orders of magnitude—from 10^{-10} to 10^{3}, the task of correctly selecting parameters *close* to the true parameters is non-trivial. In fact, without having some estimate of the actual rate parameters, it is unclear as to how to measure closeness at all. Clearly, these are among the issues that make parameter estimation so difficult for ODE and Petri net approaches. Since our comparisons will be relative and not absolute, we take a relative approach to modeling rate parameters. The space of possible rate values is *the space of possible signaling event orderings*.

This idea is illustrated in Figure 3A. Protein A affects the activity of protein D through two separate pathways. Assuming that A is active to begin with, the relative speed of these two pathways determines the final activity of D. If the pathway through C is faster than the pathway BΠD, then D will be active. However, if the pathway speeds are reversed, then D will remain inactive. The overall outcome of this network can be represented without any use of numeric reaction rates by representing the reaction rates as an ordering over all the edges in the network. We can extend this idea to the SPN by observing that there exists a unique event for each signaling edge in the signaling network.

This sampling strategy is the motivation for the dual-time framework depicted in Figure 3B and implemented by the signaling network event generator shown in Figure 6. *Time blocks* are the larger time intervals during which every signaling event occurs exactly once. Since every transition in the SPN is associated with a unique event, each transition will fire exactly once in each time block. *Transition firings* are the smaller time units that impose a strict sequential order on the occurrence of signaling events. While this strict sequentiality of firing models relative reaction rates, it also discretizes the effect of signaling events. Though this is consistent with the definition of transition firing in discrete time Petri nets (only one transition is evaluated at a given point in time) [44], in biological signaling networks there is no such serial evaluation constraint. However, our validation with experimental data suggests that this discretization approximation does not affect the overall validity of the simulation results.

PROCEDURE GENERATESIGNALINGEVENTS(E, n)

1. $k = |E|$
2. σ an an empty array of size $(k \times n)$
3. $i = 1$
4. for $b = 1$ to n
   (a) $E' = E$
   (b) while $E' \neq \emptyset$
      i. e = a random event from E'
      ii. $\sigma[i] = e$
      iii. $E' = E' - \{e\}$
      iv. $i = i + 1$
5. Return σ

Figure 6. The Algorithm That Implements the Signaling Network Event Generator. This routine generates the time block/firing structure. Given a set of events, E, and the number of blocks for which the SPN will be executed, n, GENERATESIGNALINGEVENTS generates n blocks of events, each consisting of $|E|$ events ordered randomly. In each block, every event in E occurs exactly once.
doi:10.1371/journal.pcbi.1000005.g006

Defining the Initial State

As mentioned previously, the initial state of the SPN is the initial marking, $\mathbf{m}_0$. As the SPN provides no explicit information on how this marking should be built, we propose three ways to construct the initial state: zero, basal, or experimentally derived. In a zero initial state, the simulator initializes all proteins to have zero tokens. The basal initial state is a random distribution of activation levels intended to model the cell when no impulses due directly to external stimuli are propagating through the signaling network. Though a basal network is considered at rest, in general it will not have a zero marking since signal flows are known to occur even in unstimulated signaling networks through autocrine and paracrine secretions by the cells. The experimentally derived initial state is based on knowledge about the activity levels of various proteins just prior to the addition of the external stimuli.

When accurate experimental data is available such as results from microarrays or western blots, the experimentally derived initial state may be the most accurate. A challenge in using experimental data, however, is determining how best to assign numbers of tokens based on the experimentally observed activity levels.

In the absence of reliable experimental data, the basal initial state seems more accurate than the zero initial state. However, it presents the challenge of properly selecting the basal activity-levels to assign to each protein in the model network. In [18], a basal initial state was constructed by activating a small number of randomly selected proteins in the signaling network. However, the work in [18] was done using a boolean model. Translating this approach into a tokenized model creates the additional complexity of determining how many tokens each basally active protein should receive. The correct values are likely to depend on the specific signaling network and experimental conditions.

We performed preliminary tests to compare the effect of using different basal versus zero markings on the outcome of the simulator. We found that the basal and zero states produced indistinguishable predictions so long as less than 30% of the proteins were activated and a small number of tokens (<5) were used when constructing the basal marking. This is not as surprising as it may seem at first. Inhibitory edges will quickly consume a small number of tokens scattered throughout the network, effectively returning much of the network to the zero state before a stimulation event can propagate through.

Furthermore, while validating our method, we also compared the predictions produced by SPNs based on a zero initial state and experimentally derived initial state. These, too, did not produce noticeably different final results for similar reasons as discussed above. Details of these comparisons will be discussed further in the Results and Discussion sections.

However, since all three initial state construction strategies yield qualitatively identical predictions, using zero initial states has the advantage of invoking the fewest unnecessary assumptions about the network (as in the case of the basal initial state) and requiring the least experimental data (as in the case of the experimentally derived state). Nonetheless, in our implementation of the tool, we allow for using any one of these three initial state construction strategies.

Modeling Cell-Specific Signaling Networks

Whereas consensus signaling networks typically represent the connectivity in normal cells, many experiments are conducted on abnormal cells in which oncogenic mutations, gene knockouts, and pharmacological inhibitors have altered the behavior of various signaling nodes in the network. In an SPN, these alterations to the signaling network can be modeled by adding/removing transitions

(and associated input/output arcs) and explicitly setting the token count for various proteins in the initial state.

The two network alterations which are commonly induced by oncogenic mutations, gene knockouts, or pharmacological inhibitors are constitutively high or low protein activity-levels, meaning that a protein is either unable to be inhibited or unable to be activated. The simulator allows for proteins to be specified as either fixed *High* or *Low*. Here we explain how these are modeled by changes to the SPN.

If protein u is fixed high, then this protein cannot be inhibited. Thus, all transitions that remove tokens from p_u are removed from the SPN. The fact that u is high, however, also suggests that it maintains a higher activity level in general. Therefore, in the initial state, $\mathbf{m}_0(p_u) = H$, where H is a non-zero number of tokens. Since all inhibiting transitions have been removed from the SPN, throughout any execution, place p_u will always have at least H tokens.

In experiments, we have observed that the choice of the value of H does not change the relative outcome of the simulations. While H will affect the actual number of tokens present in a given place as well as the number of time blocks required to observe certain activity-level changes, the relative changes in activity-level (number of tokens) among different proteins (places) does not change. As a result, one is free to select any reasonable value of H (for our experiments, we used H = 10) as long as this H is held constant across all simulations whose results will be compared.

If protein u is fixed low, then this protein cannot be activated. Thus, all transitions that add tokens to p_u are removed from the SPN. The fact that u is low, however, also suggests that it maintains a constantly low activity level in general. Therefore, in the initial state, $\mathbf{m}_0(p_u) = L$, where L is a small number of tokens (in our simulations we use L = 0). Since p_u is only inhibited, we observed that all constitutively low proteins quickly had their marking reduced to zero.

Unlike the value of H, extra caution must be taken when selecting values for representing L. A value of L that is too large can destabilize the early propagation of signal through the network. In our experiments, we obtained best results for values of L very close to or equal to zero ($L \leq 2$). Beyond this, the final results obtained depended on other values in the network, the strength of the signal, and the duration of the simulation.

Simulating a Signaling Network

Figure 7 provides more detailed versions of the simulation algorithm outlined in Figure 2. Steps 1 and 2 of the SIMULATE procedure constructs the initial marking and net topology to incorporate perpetually high proteins, H, and perpetually low proteins, L. In this paper, proteins that are assigned high activity-levels receive an initial token count of 10 in order to model a higher-than-average initial activity-level. As discussed earlier, using other values of H scale the activity-levels of all the proteins in the network, but will not qualitatively change their relative activities.

The loop in Step 3 runs r individual simulation runs. Each run receives a different event ordering, σ^e, thereby implementing the interaction rate sampling strategy. The time block/step structure is contained within the ordering σ^e (see Figure 6C). As a result, the SPN execution step simulates the events by firing their associated transition. Only those markings that correspond to time block boundaries are sampled.

After SIMULATE finishes collecting the time block markings from all the runs, Step 4 computes the average markings for each time block and Step 5 returns these averages.

PROCEDURE SIMULATE(S, H, L, B, r)

1. For each $p \in H$
   - $m_0(p) = 10$;
   - $I = I - \{(p,t) : t \in T \text{ and } (t,p) \notin O\}$
2. For each $p \in L$
   - $m_0(p) = 0$;
   - $I = I - \{(t,p) : t \in T\}$;
3. for $i = 1$ to r
   - $\sigma^e = \text{GenerateSignalingEvents}(E, B)$;
   - Execute $\mathbf{m}_0^i | \sigma \rangle \mathbf{m}_{B|T|}^i$;
4. For each $p \in P$ and $0 \leq b \leq B$
   - $\overline{m}_b(p) = \frac{1}{r} \sum_{i=1}^{r} m_{b|T|}^i(p)$;
5. Return $(\overline{\mathbf{m}}_1, \overline{\mathbf{m}}_2, \ldots, \overline{\mathbf{m}}_B)$

Figure 7. The Procedure for Simulating a Signaling Petri Net. SIMULATE predicts the signal flow through the SPN S. The simulation is run for B time blocks; the results of r runs are averaged to produce the final result. Most of the work is done by the signaling Petri net execution procedure detailed in the preceding sections. This execution actually performs an individual run. This procedure takes the initial marking, $\mathbf{m}_0$, and applies the sequence of transitions triggered by the event sequence, σ^e. This ordering, generated by the algorithm in Figure 6, has the dual time structure in which each block of edges contains every event in E exactly once. Each firing evaluates the effect of one transition. The markings at the end of each time block are extracted in Step 5.
doi:10.1371/journal.pcbi.1000005.g007

Simulating a Perturbation Experiment

We tested the accuracy and performance of our method by simulating the effect of two different targeted manipulations to a well-known signaling network. We compared these predictions to experimental results produced by performing the actual manipulations on two separate cancer cell lines.

The perturbations we considered in this study altered the constitutive activity-level of various proteins in the network (as opposed to affecting specific signaling interactions). Therefore, we modeled the perturbations as changes in the high and low proteins—H^c and L^c for the control (unperturbed) network and H^p and L^p for the perturbed network.

A variant of the SIMULATE method was required to quantify how a perturbation changed the protein token-counts for each time block. Figure 8 shows the algorithm we used. In the procedure DIFFERENTIALSIMULATE, the input S provides the consensus SPN. Inputs H^c and L^c specify the control high and low proteins, the inputs H^p and L^p specify the perturbed high and low proteins. After Steps 1–5 construct two separate SPNs for the control and perturbed conditions, the loop in Step 6 performs r independent simulations over the control and perturbed models. Step 6d computes the difference between the markings at the end of each time block in the perturbed and control networks. The marking difference $\boldsymbol{d}_j^i = \boldsymbol{m}_j^p - \boldsymbol{m}_j^c$ yields the marking $\boldsymbol{d}_j^i$ where $d_j^i(v) = m_j^p(v) - m_j^c(v)$ for each $v \in P$. Following the loop, the marking differences are averaged to obtain the time series $(\Delta_1, \Delta_2, \ldots, \Delta_B)$ where $\Delta_b(v)$ is the average change in the token-count for protein v at time block b.

For values of $|\Delta_b| > 0$ for a given molecule v, we can conclude that the perturbation caused a change in the activity-level of v at time block b only if the difference observed is statistically

PROCEDURE DIFFERENTIALSIMULATE(S, H^c, L^c, H^p, L^p, B, r)

1. $S^c = S$, $S^p = S$;
2. For each $p \in H^c$
   (a) $m_0^c(p) = 10$ and $I^c = I^c - \{(p,t) : t \in T^c \text{ and } (t,p) \notin O^c\}$
3. For each $p \in L^c$
   (a) $m_0^c(p) = 0$ and $I^c = I^c - \{(t,p) : t \in T^c\}$;
4. For each $v \in H^p$
   (a) $m_0^p(v) = 10$ and $I^p = I^p - \{(v,t) : t \in T^p \text{ and } (t,v) \notin O^p\}$
5. For each $v \in L^p$
   (a) $m_0^p(v) = 0$ and $I^p = I^p - \{(t,v) : t \in T^p\}$;
6. for $i = 1$ to r
   (a) $\sigma^e = \text{GenerateSignalingEvents}(E,T)$;
   (b) Execute $\mathbf{m}_0^c[\sigma\rangle\mathbf{m}_{B|T|}^c$;
   (c) Execute $\mathbf{m}_0^p[\sigma\rangle\mathbf{m}_{B|T|}^p$;
   (d) For $j = 0$ to B
      i. $\mathbf{d}_j^i = \mathbf{m}_{j|T|}^p - \mathbf{m}_{j|T|}^c$
7. For each $p \in P$ and $0 \le b \le B$
   (a) $\Delta_b(p) = \frac{1}{r}\sum_{i=1}^{r} d_b^i(p)$;
8. Return $(\Delta_1, \Delta_2, \ldots, \Delta_B)$;

Figure 8. The Algorithm for Predicting the Effect on Signal Propagation of a Targeted Manipulation. The algorithm for predicting the effect on signal propagation of a targeted manipulation on signaling network with connectivity G. The 'c' and 'p' superscripts are used to denote parameters in the *control* and *perturbed* versions, respectively, of the SPN.
doi:10.1371/journal.pcbi.1000005.g008

significant. We use a t-test to determine whether this change is statistically significant for protein v at time block b. Computing the t-test for two distributions (control and perturbation) requires knowledge of the mean ($\mu_{c,b}$ and $\mu_{p,b}$) as well as the variance $\left(\sigma_c^2 \text{ and } \sigma_p^2\right)$ for both distributions. In order to obtain these parameters for the control network, a large number, X, of independent simulations is run. Simulation i provides a single series of markings, $\left(\overline{\boldsymbol{m}}_1^i, \overline{\boldsymbol{m}}_2^i, \ldots, \overline{\boldsymbol{m}}_B^i\right)$. The mean is then computed:

$$\mu_{c,b,v} = \frac{\sum_{i=1}^{X} \overline{\boldsymbol{m}}_{b(v)}^i}{X}.$$

The variance is computed similarly:

$$\sigma_{c,b,v}^2 = \frac{\sum_{i=1}^{X} \left(\overline{\boldsymbol{m}}_{b(v)}^i - \mu_{c,b,v}\right)^2}{X-1}.$$

The parameters $\mu_{p,b,v}$ and $\sigma_{p,b,v}^2$ for the perturbed network are computed as described above by substituting the perturbed network for the control network. Using these parameters, the t-value for molecule v at time block b can be computed from the formula

$$t-value = \frac{\mu_{c,b,v} - \mu_{p,b,v}}{\sqrt{\frac{\sigma_{c,b,v}^2}{X} + \frac{\sigma_{p,b,v}^2}{X}}}.$$

The statistical significance of the difference can then be obtained by comparing the t-value to the desired critical value.

Note that the DIFFERENTIALSIMULATE procedure and the associated significance test can predict the effect not only of perturbations, but also of any two different experimental (or cellular) conditions imposed on the same signaling network. As a result, in addition to perturbation experiments, our method can also be used to study the effects of other phenomena that induce changes in the propagation of signal through a signaling network.

Cell-Specific Signaling Network Models

Figure 1 shows the signaling network we analyzed. We obtained the core connectivity from a published literature survey on the EGFR network [57]. We added to this several other well-established interactions taken from literature [36–43]. The response of this network to various perturbations was measured and simulated in two separate breast cancer cell lines: MDA231 and BT549. The core signaling Petri net used, S^{EGFR}, is captured by the following signaling proteins and interactions: places (the set P): v_{EGFR}, v_{SRC}, v_{Rac}, v_{MEKK4}, v_{MEK4}, v_{JNK}, v_{MEKK6}, v_{MEK6}, v_{STAT}, v_{Grb2}, v_{Shc}, v_{SOS}, v_{RB}, v_{ELK}, v_{BAD}, v_{NFKB}, v_{RAS}, v_{GAB1}, v_{PIP3}, v_{PI3K}, v_{PDK1}, v_{PTEN}, $v_{c\text{-}Raf}$, v_{AKT},v_{LKB1}, v_{MEK}, $v_{GSK3\beta}$, v_{AMPK}, v_{TSC2}, $v_{MAPK1,2}$, v_{RSK}, v_{Rheb}, $v_{mTOR\text{-}Raptor}$, v_{4EBP1}, v_{p70S6K}, v_{p38}, and v_{pS6}.

Protein interaction network motifs (the combination of arcs and transitions): $v_{EGFR} \rightarrow v_{Grb2}$, $v_{Grb2} \rightarrow v_{Shc}$, $v_{Shc} \rightarrow v_{SOS}$, $v_{SOS} \rightarrow v_{RAS}$, $v_{Grb2} \rightarrow v_{GAB1}$, $v_{GAB1} \rightarrow v_{PI3K}$, $v_{EGFR} \rightarrow v_{SRC}$, $v_{SRC} \rightarrow v_{STAT}$, $v_{PI3K} \rightarrow v_{PIP3}$, $v_{PIP3} \rightarrow v_{PDK1}$, $v_{RAS} \rightarrow v_{c\text{-}Raf}$, $v_{PDK1} \rightarrow v_{AKT}$, $v_{RAS} \rightarrow v_{Rac}$, $v_{Rac} \rightarrow v_{MEKK4}$, $v_{MEKK4} \rightarrow v_{MEK4}$, $v_{MEK4} \rightarrow v_{JNK}$, $v_{JNK} \rightarrow v_{STAT}$, $v_{Rac} \rightarrow v_{MEKK6}$, $v_{MEKK6} \rightarrow v_{MEK6}$, $v_{MEK6} \rightarrow v_{p38}$, $v_{p38} \rightarrow v_{STAT}$, $v_{PDK1} \rightarrow v_{p70S6K}$, $v_{PTEN} \dashv v_{AKT}$, $v_{AKT} \dashv v_{c\text{-}Raf}$, $v_{AKT} \dashv v_{GSK3\beta}$, $v_{AKT} \dashv v_{TSC2}$, $v_{AKT} \dashv v_{AMPK}$, $v_{AKT} \dashv v_{BAD}$, $v_{AKT} \rightarrow v_{NFKB}$, $v_{AKT} \rightarrow v_{p70S6K}$, $v_{LKB1} \rightarrow v_{AMPK}$, $v_{MEK} \rightarrow v_{MAPK1,2}$, $v_{MAPK1,2} \rightarrow v_{RB}$, $v_{MAPK1,2} \rightarrow v_{ELK}$, $v_{MAPK1,2} \rightarrow v_{STAT}$, $v_{GSK3\beta} \rightarrow v_{TSC2}$, $v_{AMPK} \rightarrow v_{TSC2}$, $v_{MAPK1,2} \dashv v_{EGFR}$, $v_{MAPK1,2} \dashv v_{TSC2}$, $v_{MAPK1,2} \rightarrow v_{p70S6K}$, $v_{MAPK1,2} \rightarrow v_{RSK}$, $v_{RSK} \dashv v_{TSC2}$, $v_{TSC2} \dashv v_{Rheb}$, $v_{Rheb} \rightarrow v_{mTOR\text{-}Raptor}$, $v_{AKT} \rightarrow v_{mTOR\text{-}Raptor}$, $v_{mTOR\text{-}Raptor} \rightarrow v_{4EBP1}$, $v_{mTOR\text{-}Raptor} \rightarrow v_{p70S6K}$, $v_{p70S6K} \dashv v_{EGFR}$, $v_{SRC} \dashv v_{SRC}$, $v_{Rac} \dashv v_{Rac}$, $v_{MEKK4} \dashv v_{MEKK4}$, $v_{MEK4} \dashv v_{MEK4}$, $v_{JNK} \dashv v_{JNK}$, $v_{MEKK6} \dashv v_{MEKK6}$, $v_{MEK6} \dashv v_{MEK6}$, $v_{STAT} \dashv v_{STAT}$, $v_{Grb2} \dashv v_{Grb2}$, $v_{Shc} \dashv v_{Shc}$, $v_{SOS} \dashv v_{SOS}$, $v_{RAS} \dashv v_{RAS}$, $v_{c\text{-}Raf} \dashv v_{c\text{-}Raf}$, $v_{MEK} \dashv v_{MEK}$, $v_{MAPK1,2} \dashv v_{MAPK1,2}$, $v_{RB} \dashv v_{RB}$, $v_{ELK} \dashv v_{ELK}$, $v_{RSK} \dashv v_{RSK}$, $v_{GAB1} \dashv v_{GAB1}$, $v_{PIP3} \dashv v_{PIP3}$, $v_{p38} \dashv v_{p38}$, $v_{PI3K} \dashv v_{PI3K}$, $v_{PDK1} \dashv v_{PDK1}$, $v_{AKT} \dashv v_{AKT}$, $v_{BAD} \dashv v_{BAD}$, $v_{NFKB} \dashv v_{NFKB}$, $v_{AMPK} \dashv v_{AMPK}$, $v_{mTOR\text{-}Raptor} \dashv v_{mTOR\text{-}Raptor}$, $v_{p70S6K} \dashv v_{p70S6K}$, $v_{pS6} \dashv v_{pS6}$, $v_{4EBP1} \dashv v_{4EBP1}$.

Notice that the last several edges are self-inhibitory loops (e.g., $v_{Ras} \dashv v_{Ras}$). These loops are used to model regulatory mechanisms that are not present in the model network.

For molecules that do not have specific inhibitory edges modeled in the network, we use the self-inhibitory loop to prevent exponential increase in the token counts and to model inhibitory mechanisms beyond the scope of the network. For example, consider the molecule Ras in the network shown in Figure 1. In the model, this protein is not inhibited. However, biologically we know that Ras has intrinsic GTPase function which inactivate itself. In order to model this, we introduce a self-inhibitory loop.

The differences between the two cell-specific networks are captured by following activity assignments to various proteins in the SPN. In the MDA231 cell line, $H^{MB} = \{v_{Ras}, v_{EGF}\}$ and $L^{MB} = \emptyset$. In the BT549 cell line, $H^{BT} = \{v_{EGF}\}$ and $L^{BT} = \{v_{PTEN}\}$.

Of the two perturbations we considered, one significantly knocked down the activity-level of TSC2 and the other knocked down mTOR-Raptor. While the core SPN still modeled these networks, separate *perturbed* activity-assignments were required for each cell line-perturbation pairing: $L^{MB\text{-}TSC2} = L^{MB} \cup \{v_{TSC2}\}$,

$L^{MB\text{-}mTOR} = L^{MB} \cup \{v_{mTOR\text{-}Raptor}\}$, $L^{BT\text{-}TSC2} = L^{BT} \cup \{v_{TSC2}\}$ and $L^{BT\text{-}mTOR} = L^{BT} \cup \{v_{mTOR\text{-}Raptor}\}$.

Setup for Perturbation Experiments

Cell culture and stimulation. Human MDA-MB-231 (MDA231) and BT549 breast cancer cells were routinely maintained in RPMI supplemented with 10% FBS. For signaling experiments, logarithmically growing cells were serum-starved for 16 hours and then subjected to treatments by epidermal growth factor (EGF) (20 ng/mL) (Cell Signaling Technology, Beverly, Massachusetts) for 30 minutes. Controls were incubated for corresponding times with DMSO. To knock down TSC2, cells were treated with short interfering RNA (siRNA) (Dharmacon, Lafayette, Colorado) for 72 hours prior to EGF stimulation. Control cells were transfected with non-targeting (N/T) siRNA (Dharmacon, Lafayette, Colorado) prior to EGF treatment.

Antibodies. The following antibodies were used for immunoblotting: anti-phospho-p44/42 MAPK, anti-phospho-GSK3β (S21/S9); anti-phospho-AKT(ser473); anti-phospho-TSC2(T1462); anti-phospho-mTOR(S2448); anti-phospho-P70S6K(T389) (Cell Signaling Technology, Boston, Massachusetts); and anti-β-Actin (Sigma-Aldrich, St. Louis, Missouri).

SDS-PAGE and immunoblotting. Cells were lysed by incubation on ice for 15 minutes in a sample lysis buffer (50 mM Hepes, 150 mM NaCl, 1 mM EGTA, 10 mM Sodium Pyrophosphate, pH 7.4, 100 nM NaF, 1.5 mM MgCl2, 10% glycerol, 1% Triton X-100 plus protease inhibitors; aprotinin, bestatin, leupeptin, E-64, and pepstatin A). Cell lysates were centrifuged at 15,000 g for 20 minutes at 4°C. The supernatant was frozen and stored at −20°C. Protein concentrations were determined using a protein-assay system (BCA, Bio-Rad, Hercules, California), with BSA as a standard. For immunoblotting, proteins (25 μg) were separated by SDS-PAGE and transferred to Hybond-C membrane (GE Healthcare, Piscataway, New Jersey). Blots were blocked for 60 minutes and incubated with primary antibodies overnight, followed by goat anti-mouse IgG-HRP (1:30,000; Cell Signaling Technology, Boston, Massachusetts) or goat anti-rabbit IgG-HRP (1:10,000; Cell Signaling Technology) for 1 hour. Secondary antibodies were detected by enhanced chemiluminescence (ECL) reagent (GE Healthcare, Piscataway, New Jersey). All experiments were repeated a minimum of three independent times.

Setup for perturbation simulations. To select the block duration parameter, B, we compared the experimentally derived fold change of AKT in the MDA231 cell line to the AKT fold changes predicted for B = 10, 20, 50, 100, and 1000. We found B = 20 to be the best fit and used this value for all simulations in this study.

We also experimented with input parameter r, the numbers of individual simulation runs averaged per simulation. We tried a range extending from $r = 100$ to $r = 1000$. We found that no observable changes occurred in trends for $r \geq 400$. Therefore, $r = 400$ was used for all simulations in this study.

We considered both the zero and experimentally derived initial states as the initial markings for the TSC inhibition simulations. The experimental states for both cell lines were derived from western blots produced from cells that were incubated in DMSO and serum-starved for 16 hours. Unsampled molecules were assigned a marking of zero. The number of tokens assigned to each sampled molecule was directly proportional to the darkness of the line on the western blot. This assignment was done by hand, though devising automated and standardized methods for the construction of experimentally derived initial states is an important direction for future work. Since most of the molecules in the network were not sampled, only mTOR-Raptor, TSC2, GSK3β, p70S6K, AKT, and MAPK were given non-zero markings. The initial markings used are shown in Table 1.

Since experimental results for the mTOR-Raptor inhibition were obtained from literature, we did not have experimental results for construction of experimentally derived initial states. Therefore, we used the zero initial states for the mTOR-Raptor inhibition simulations.

Results

In order to evaluate the accuracy of our simulation method, we tested its predictions of the effect of targeted manipulations on two cell-specific versions of the signaling network depicted in Figure 1. In each cell line, a TSC2-specific siRNA was applied and the concentration of several key proteins in the EGFR network were sampled 30 minutes after stimulation with EGF. This was repeated in the absence of the TSC2 siRNA in order to obtain the concentration in the control network. We also collected a corpus of literature detailing the response of signaling proteins activity-levels to the inhibition of mTOR-Raptor using Rapamycin [43,58]. Predictions were generated by our simulator for the TSC2 and mTOR-Raptor perturbations in both cell lines.

Simulation

To simulate a perturbation, we used two networks both based on the signaling network shown in Figure 1: the control network for the cell line and the perturbed network for the cell line. The control networks for the cell lines were different because it was important to model the cell-specific mutations. In the case of the BT549 cell line, there is a mutation that leads to the loss of PTEN, which makes AKT always active. In the MDA231 cell line, there is a mutation in Ras, which makes it always active. As shown in the formulation of the model, these are modeled using fixed activity assignments in the simulator.

The TSC2 (mTOR-Raptor) perturbed network for a cell line was created by taking the control network and fixing the activity-level of TSC2 (mTOR-Raptor) to zero for the duration of the simulation, effectively simulating the pharmacological inhibition of the protein. For each cell-line/perturbation pair, we ran the simulator on the control and perturbed networks using the DifferentialSimulate procedure in Figure 8 which computed the change in token-counts induced by the perturbation for all proteins in the model. These change plots are shown in Figure 9 for TSC2 and in Figure 10 for mTOR-Raptor. We ran the simulations using both experimentally derived initial states as well

Table 1. Experimentally Derived Initial Markings Used in the Simulations.

Molecule	*MB231*		*BT549*	
	Control	TSC2 Inhibited	Control	TSC2 Inhibited
mTOR-Raptor	0	1	5	5
TSC2	0	0	6	0
GSK3β	5	3	3	6
p70S6K	0	2	0	0
AKT	0	0	7	7
MAPK	2	6	1	2

doi:10.1371/journal.pcbi.1000005.t001

as zero initial states. The initial state used did not change the overall trends observed in the simulations.

Using the t-test described in the Methods section, we also computed the statistical significance of the final time block (b = 20) for each molecule considered. For each molecule considered, 400 runs, 20 time blocks, and 50 samples were used. With the exception of GSK3β which did not show a significant response to the perturbation, the changes of all other proteins sampled were beyond the 0.05 significance level (see Table 2). The statistical insignificance of the change in GSK3β is not surprising since, as shown in Figure 1, GSK3β is solely activated by LKB, a molecule fixed high in both cell lines. Thus, we should not expect either perturbation to have a significant effect on the activity of GSK3β, which is what the t-value indicates.

Experimental Results

After the TSC2 perturbation was applied to a cell line, the protein concentrations were collected using western blots. Details are given in the Materials and Methods section. The western blot results are shown in Figure 9.

Discussion

As can be seen in Table 3, our method correctly predicted the *relative* protein activity-level changes induced by the TSC2 perturbation in both cell lines, for most molecules sampled. Notice that *no change* (–) was reported for the predicted response of MAPK to the TSC2 perturbation despite the fact that a small change did occur in its marking during the simulation (see Figure 9)

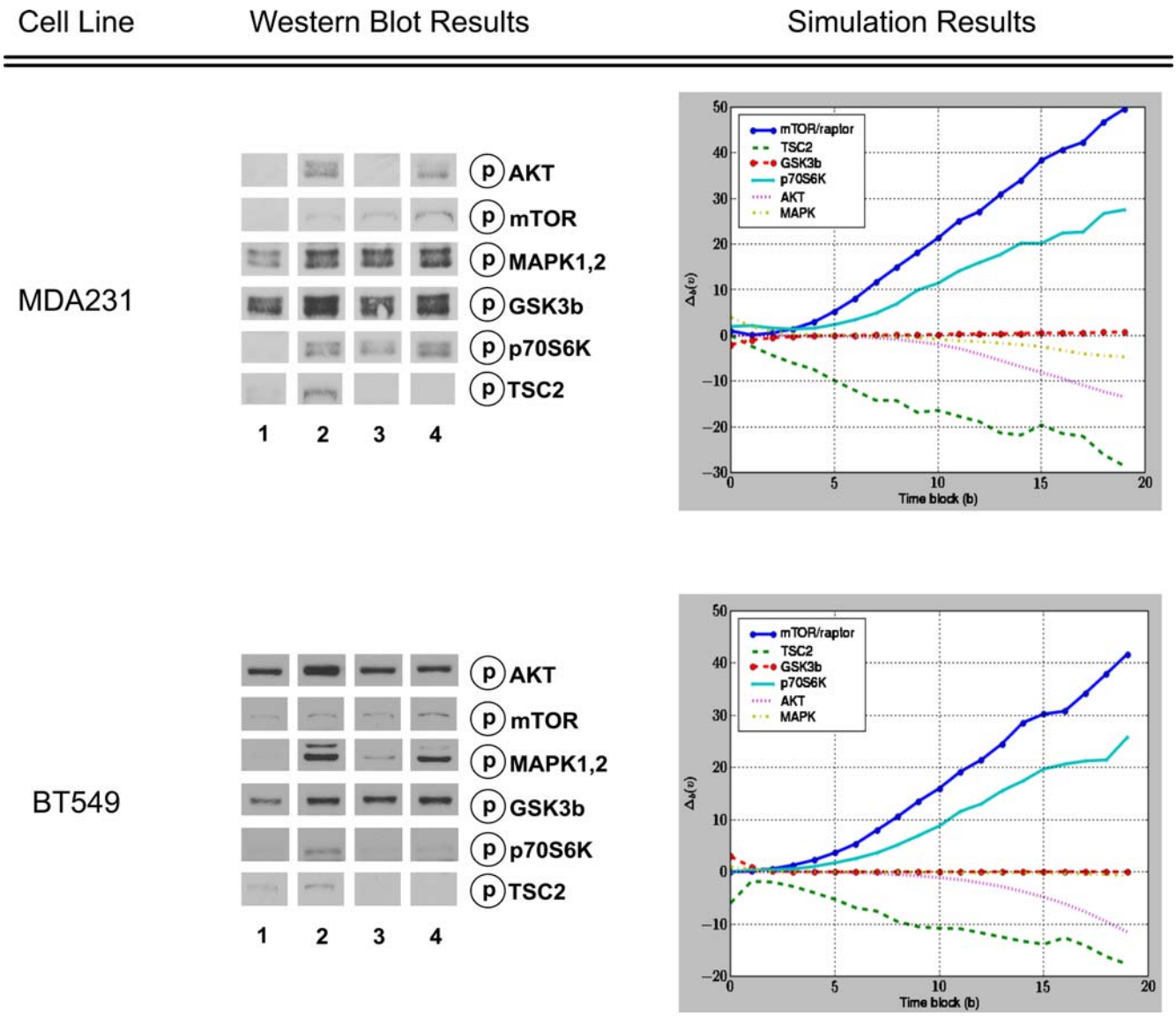


Figure 9. The Results of the TSC2 Perturbation Experiments and Simulations. In the western blots, columns (or lanes) are as follows: (1) non-targeting (NT) control siRNA, (2) NT siRNA+EGF, (3) TSC2 siRNA, (4) TSC2 siRNA+EGF. The effect of the TSC2 siRNA on a given molecule can be assessed by comparing column 4 against column 2. For each molecule in the western blot, there is a corresponding simulation curve showing the predicted change in protein activity over time. For the purposes of this analysis, we compared the concentration change after 20 time steps (the left-most data points in the plots) for each molecule. Each simulation point corresponds to the average of 400 measurements that were computed using the procedure described in Figure 8. Experimentally derived initial states were used in the simulations. The results of both the experiments and simulations are qualitatively summarized in Table 3.
doi:10.1371/journal.pcbi.1000005.g009

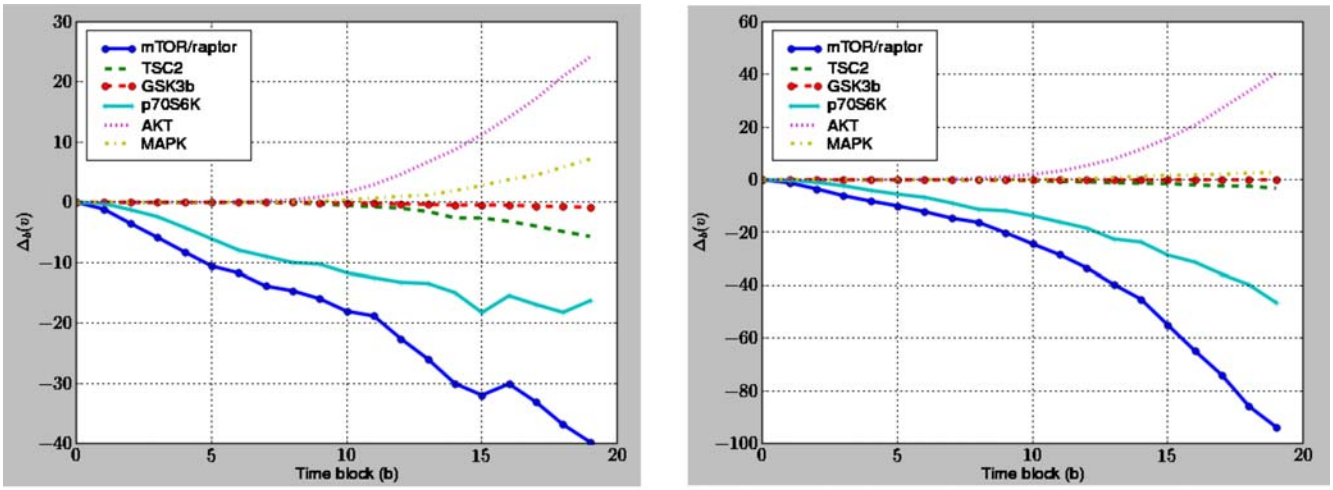


Figure 10. The Predicted Response of the Network to an mTOR-Raptor Perturbation. The predicted response of the network to a mTOR-Raptor perturbation in the (A) MDA231 and (B) BT549 cell lines. Our method predicts that the amount of available AKT increases in response to the perturbation, which is in agreement with results published in the literature [43,58]. Our method also predicts that the activity-level of p70S6K in the MDA231 cell line decreases in response to the perturbation, which has been observed experimentally [59]. Each point corresponds to the average of 400 measurements that were computed using the procedure described in Figure 8.
doi:10.1371/journal.pcbi.1000005.g010

and the t-value for the change is significant (see Table 2). At first, interpreting this value as *no change* may seem misleading. However, one of the significant challenges in experimental perturbation experiments is separating true system responses from the background noise created by experimental variables that cannot be precisely controlled (among them cell population sizes, variability in microarray antibody binding effectiveness, and limited sensitivity of hardware and software used to quantify experimental results). As a result, a common practice is to only consider those substantial changes that are well beyond the background noise level. Our interpretation of the small predicted change in MAPK as *no change* reflects the fact that such small changes would not be detectable in microarray or western blot results. Thus, though such a small fluctuation might have occurred in the real data, it would not have been detected by the biologists and most likely would appear in the experimental data to have not changed.

Similar reasoning guided our decision to characterize the simulation (and experimental) results as either up (↑), down (↓), or no change (−) in general. Since the amount of protein registered in a microarray or western blot is not always a reliable indicator of the exact amount of protein (or protein form) being measured, biologists are often reluctant to report degrees of increases or decreases—preferring binary observations such as *up* or *down* which are less subject to influence by extraneous experimental conditions. It is true that our simulation method produces precisely quantified increases or decreases which can be taken to indicate degrees of change in response to perturbations. However, as experimental techniques cannot reliably measure degrees of increase or decrease, we judged the binary (up or down) characterization to be a more reliable way of validating our method. Certainly, our method provides additional information of

Table 2. The T-Values for the Molecules Sampled in the Microarray.

Molecule	t-Value in MDA231	t-Value in BT549
mTOR-Raptor	41.72	30.53
TSC2	21.65	8.28
GSK3β	0.42	0.10
p70S6K	14.22	5.83
AKT	6.60	9.55
MAPK	16.35	18.93

The critical value for an alpha value of 0.05 with 50 samples is 2.0086. Note that the t-values for all molecules except for GSK3β are larger than this value, confirming that these changes are statistically significantly.
doi:10.1371/journal.pcbi.1000005.t002

Table 3. Summary of the Effect of Perturbation Reported by Experimental and Simulated Methods.

Molecule	MB231		BT549	
	Experiment	Simulation	Experiment	Simulation
mTOR-Raptor	↑	↑	↑ or −	↑
TSC2	↓	↓	↓	↓
GSK3β	−	−	−	−
p70S6K	↑	↑	↓	↑
AKT	↓ or −	↓	↓	↓
MAPK	−	−	−	−

The up arrow (↑) indicates that the perturbation caused a rise in the level of the phosphorylated protein; the straight line (−) indicates no change; and the down arrow (↓) indicates that a decrease occurred. Values in the *Experiment* column were estimated by comparing lanes 4 and 2 in Figure 9. We estimated the *Simulation* column by determining whether the top quartile of the distribution for the final time point was above, below, or at zero. In some cases it is difficult to judge for certain whether the total quantity of the phosphorylated protein changed or remained the same—both for the experimental and computational cases. In these situations, we indicated the uncertainty by listing the possible changes that the protein *could* have feasibly undergone.
doi:10.1371/journal.pcbi.1000005.t003

degrees of change and we consider studying the accuracy of these degrees to be an important area for future work.

Our method also correctly predicted the activity-level change of AKT in response to mTOR-Raptor inhibition as reported by a number of studies [43,58]. Further, our method predicted that, when mTOR-Raptor is inhibited, the level of p70S6K in the MDA231 cell line decreased, which also had been observed experimentally [59].

The only incorrect prediction made by our method was the activity-level change of p70S6K in the BT549 cell line. However, BT549 cells contain an RB mutation [49] which could alter p70S6K phosphorylation [60]. It is a strength of our simulator that the discrepancy between our method's predictions and the experimental results identified a section of the model in which additional connectivity has been found which might account for the difference observed.

The predictions made by our simulator would be exceedingly difficult to derive by visual or manual inspection. Table 4 shows the number of paths between several pairs of compounds within the network. Where there is more than one path connecting two molecules, feed forward and feed backward loops are present. Attempting to determine, by hand, how these different loops will interact with one another is, by itself, a difficult endeavor even when not considering the additional task of deriving the rest of the network dynamics simultaneously. For the larger networks that are now becoming available, computational analysis becomes even more crucial to obtaining insights into the dynamic behavior of the network.

Despite the complexity of the network dynamics, it was straightforward to find and integrate the connectivity information used to build it. Most of the information sources [36–43] established the *existence* of various pathways and provided few or no biochemical or kinetic details. As a result, the literature we used would have provided little assistance is building a parameterized Petri net or ODE model. Due to the proliferation of curated signaling network repositories and searchable literature archives, connectivity information is relatively abundant which makes the ad hoc assembly of networks a relatively straightforward endeavor. This further underscores the advantage of using our method over ODEs or parameterized Petri nets to quickly model and characterize some of the dynamics of a signaling network.

For simulations that will be compared to experimental results, the time parameter must be selected carefully. The time parameter, B, indicates how many time blocks our method will simulate. The time block is an abstract unit of time. Therefore, before comparing experimental results and predictions, it is necessary to determine how many seconds, minutes, or hours correspond to a time block. This can be done by comparing a prediction of the simulator with the experimentally measured activity-level of one or two proteins at several time points in order to determine what time blocks correspond to the different sampled time points. In the present study, we calibrated our time blocks only once for two cell lines and six experimental conditions (two cell lines, with/without TSC2, with/without mTOR-Raptor). To select the time parameter we used the experimentally measured activity changes in two proteins at two time points. In contrast to other predictive dynamic analysis tools which require multiple time points and multiple protein samples in order to calibrate simulation and model parameters, our method has relatively low time and resource investment.

Besides the time parameter, the other component of our simulations which involved experimentally obtained knowledge was the initial states. The experimentally derived initial states require that some experimental data be available providing information on the initial concentrations of individual signaling proteins in the network prior to stimulation. However, in the network that we considered here, the overall behavior of the network and of individual signaling proteins was resilient to changes in the initial states used. Zero and experimentally derived both produced the same overall change predictions. Thus, while experimentally derived initial states may be important for the simulation of some networks, it may well be the case that many networks (such as the one we considered in this paper) can be simulated without this knowledge—further reducing the experimental work that must be done prior to simulation.

Table 4. Number of Paths Connecting Several Pairs of Compounds in the EGFR Model Used in Our Simulations

Source Protein	Destination Protein	Number of Paths
EGFR	TSC2	7
AKT	mTOR-Raptor	6
MEK	EGFR	4
AKT	p70S6K	8

The multiple paths connecting pairs of proteins highlight the complex interactions present within the network that give rise to its overall dynamic behavior.

doi:10.1371/journal.pcbi.1000005.t004

The fact that our simulator produced accurate predictions for a variety of experimental conditions using the one core network model and set of simulation parameters also distinguishes our method from other predictive approaches. The only aspects of the model that were modified during the simulations were activity-levels reflecting the immediate effects of either the underlying tumor mutations (Ras and PTEN) or the perturbations (mTOR-Raptor and TSC2 targeted manipulation). In contrast, the accuracy of ODEs and Petri nets predictions are known to be sensitive to small changes to the model. For comparative studies such as the one conducted in this paper, an ODE or parameterized Petri net model might need to be re-constructed with different parameters for each experimental condition of interest. As a result, while it is possible to obtain our simulation results using these models, it remains beyond the capabilities of any existing ODE or parameterized Petri net system to provide insights into the effects of experimental conditions on the dynamic behavior of a signaling network with so little initial time and resource investment.

Though our method's predictions will not be as accurate as the results returned by a correctly parameterized ODE, biologists using our method can derive information about a network's dynamic behavior without having to conduct extensive experimentation and computationally expensive parameter estimation. This novel capability offers scientists the exciting prospect of being able to test hypotheses regarding signal propagation in silico. As a result, by using our method researchers can evaluate a wide array of network responses in order to determine the most promising experiments before even entering the laboratory.

Acknowledgments

We gratefully acknowledge the three anonymous reviewers, whose comments and feedback improved the manuscript significantly.

Author Contributions

Conceived and designed the experiments: DR MM JTT PTR. Performed the experiments: DR MM JTT. Analyzed the data: DR MM LN PTR. Contributed reagents/materials/analysis tools: DR MM LN PTR. Wrote the paper: DR MM LN PTR.

References

1. Hunter T (2000) Signaling-2000 and beyond. Cell 100: 113–127.
2. Hanahan D, Weinberg RA (2000) The Hallmarks of Cancer. Cell 100: 57–70.
3. Feldman DS, Carnes CA, Abraham WT, Bristow MR (2005) Mechanisms of Disease: beta-adrenergic receptors alterations in signal transduction and pharmacogenomics in heart failure. Nature Clinical Practice Cardiovascular Medicine 2: 475–483.
4. Belloni E, Muenke M, Roessier E, Traverse G, Siegel-Bartelt J, et al. (1996) Identification of Sonic hedgehog as a candidate gene responsible for holoprosencephaly. Nat Genet 14: 353–356.
5. Ma'ayan A, Jenkins SL, Neves S, Hasseldine A, Grace E, et al. (2005) Formation of regulatory patterns during signal propagation in a Mammalian cellular network. Science 309: 1078.
6. Papin JA, Hunter T, Palsson BO, Subramaniam S (2005) Reconstruction of cellular signalling networks and analysis of their properties. Nat Rev Mol Cell Biol 6: 99–111.
7. The Cancer Cell Map (2006) http://cancer.cellmap.org/cellmap/.
8. Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 28: 27–30.
9. Thomas PD, Campbell MJ, Kejariwal A, Mi J, Karlak B, et al. (2003) PANTHER: a library of protein families and subfamilies indexed by function. Genome Res 13: 2129–2141.
10. Dasika MS, Burgard A, Maranas CD (2006) A computational framework for the topological analysis and targeted disruption of signal transduction networks. Biophysical J 91: 382–398.
11. Eker S, Knapp M, Laderoute K, Lincoln P, Talcott C (2002) Pathway Logic: Executable Models of Biological Networks. Electronic Notes Theoretical Computer Science 71.
12. Ruths D, Nakhleh L, Iyengar MS, Reddy SAG, Ram PT (2006) Graph-theoretic Hypothesis Generation in Biological Signaling Networks. J Computational Biology 13: 1546–1557.
13. Schaub MA, Henzinger TA, Fisher J (2007) Qualitative networks: a symbolic approach to analyze biological signaling networks. BMC Systems Biology 1: 4.
14. Papin JA, Palsson BO (2004) The JAK-STAT signaling network in the human B-cell: an extreme signaling pathway analysis. Biophysical J 87: 37–46.
15. Papin JA, Price ND, Wiback SJ, Fell DA, Palsson BO (2003) Metabolic pathways in the post-genomic era. Trends Biochemical Sciences 28: 250–258.
16. Schilling CH, Letscher D, Palsson BO (2000) Theory for the Systemic Definition of Metabolic Pathways and their use in Interpreting Metabolic Function from a Pathway-Oriented Perspective. J Theoretical Biology 203: 229–248.
17. Chaouiya C (2007) Petri net modelling of biological networks. Briefings Bioinformatics 8: 210–219.
18. Li C, Suzuki S, Nakata M (2006) Structural Modeling and Analysis of Signaling Pathways Based on Petri Nets. J Bioinformatics Computational Biology 4: 1119–1140.
19. Sackmann A, Heiner M, Koch I (2006) Application of Petri net based analysis techniques to signal transduction pathways. BMC Bioinformatics 7: 482–498.
20. Steggles LJ, Banks R, Shaw O, Wipat A (2007) Qualitatively modelling and analysing gene regulatory networks: a Petri net approach. Bioinformatics 23: 336–343.
21. Bhalla US, Ram PT, Iyengar R (2002) MAP kinase phosphatase as the locus of flexibility in a mitogen-activated protein kinase signaling network. Science 297: 1018–1023.
22. Bornheimer SJ, Maurya MR, Farquhar MG, Subramaniam S (2004) Computational modeling reveals how interplay between components of a GTPase-cycle module regulates signal transduction. Proc Natl Acad Sci U S A 101: 15899–15904.
23. Hoffman A, Levchenko A, Scott ML, Baltimore D (2002) The IkB-NF-kB signaling module: temporal control and selective gene activation. Science 298: 1242–1245.
24. Huang CY, Ferrell JE (1996) Ultrasensitivity in the mitogen-activated protein kinase cascade. Proc Natl Acad Sci U S A 93: 10078–10083.
25. Ferrell JE, Machleder EM (1998) The biochemical basis of an all-or-none cell fate switch in Xenopus oocytes. Science 280: 895–898.
26. Bailey JE (2001) Complex biology with no parameters. Nat Biotechnol 19: 503–504.
27. Novere NL, Finney A, Hucka M, Bhalla US, Campagne F, et al. (2005) Minimum information requested in the annotation of biochemical models (MIRIAM). Nat Biotechnol 23: 1509–1515.
28. Arisi I, Cattaneo A, Rosato V (2006) Parameter estimate of signal transduction pathways. BMC Neuroscience 7: S6.
29. Doi A, Fujita S, Matsuno H, Nagasaki M, Miyano S (2004) Constructing Biological Pathway Models with Hybrid Functional Petri Nets. In Silico Biology 4: 271–291.
30. Hardy S, Robillard PN (2004) Modeling and Simulation of Molecular Biology Systems using Petri Nets: Modeling Goals of Various Approaches. J Bioinformatics Computational Biology 2: 619–637.
31. Matsuno H, Tanaka Y, Aoshima H, Doi A, Matsui M, et al. (2003) Biopathways representation and simulation on hybrid functional Petri net. In Silico Biology 3: 389–404.
32. Li S, Assmann SM, Albert R (2006) Predicting essential components of signal transduction networks: a dynamic model of guard cell abscisic acid signaling. PLoS Biology 4: e312–e328.
33. Glass L, Kauffman A (1973) The logical analysis of continuous non-linear biochemical control networks. J Theoretical Biology 39: 103–129.
34. Jong HD, Geiselmann J, Batt G, Hernandez C, Page M (2004) Qualitative Simulation of the Initiation of Sporulation in Bacillus subtilis. Bull Mathematical Biology 66: 261–299.
35. Muller M, Obeyesekere M, Mills GM, Ram PT (2007) Network topology determines dynamics of the mammalian MAPK1,2 signaling network: bi-fan motif regulation of C-Raf and B-Raf isoforms by FGFR and MC1R. FASEB J; In press.
36. Avruch J, Hara K, Lin Y, Liu M, Long X, et al. (2006) Insulin and amino-acid regulation of mTOR signaling and kinase activity through the Rheb GTPase. Oncogene 25: 6361–6372.
37. Inoki K, Ouyang H, Zhu T, Lindvall C, Wang Y, et al. (2006) TSC2 integrates Wnt and energy signals via a coordinated phosphorylation by AMPK and GSK3 to regulate cell growth. Cell 126: 955–968.
38. Karbowniczek M, Cash T, Cheung M, Robertson GP, Astrinidis A, et al. (2004) Regulation of B-Raf kinase activity by tuberin and Rheb is mTOR-independent. J Biological Chemistry 279: 29930–29937.
39. Kwiatkowski DJ, Manning BD (2005) Tuberous sclerosis: a GAP at the crossroads of multiple signaling pathways. Human Molecular Genetics 14: R251–R258.
40. Liang J, Shao SH, Xu ZX, Hennessy B, Ding Z, et al. (2007) The energy sensing LKB1-AMPK pathway regulated p27(kip1) phosphorylation mediating the decision to enter autophagy or apoptosis. Nature Cell Biology 9: 218–224.
41. Ma L, Chen Z, Erdjument-Bromage H, Tempst P, Pandolfi PP (2005) Phosphorylation and functional inactivation of TSC2 by Erk implications for tuberous sclerosis and cancer pathogenesis. Cell 121: 179–193.
42. Manning BD, Logsdon MN, Lipovsky AI, Abbott D, Kwiatkowski DJ, et al. (2005) Feedback inhibition of Akt signaling limits the growth of tumors lacking Tsc2. Genes and Development 19: 1773–1778.
43. O'Reilly KE, Rojo F, She QB, Solit D, Mills GB, et al. (2006) mTOR inhibition induces upstream receptor tyrosine kinase signaling and activates Akt. Cancer Research 66: 1500–1508.
44. David R, Alla H (2005) Discrete, Continuous, and Hybrid Petri Nets. Springer.
45. Aldana M, Cluzel P (2003) A natural class of robust networks. Proc Natl Acad Sci 100: 8710–8714.
46. Kauffman S, Peterson C, Samuelsson B, Troein C (2004) Genetic networks with canalyzing Boolean rules are always stable. Proc Natl Acad Sci 101: 17102–17107.
47. Klemm K, Bornholdt S (2005) Topology of biological networks and reliability of information processing. Proc Natl Acad Sci U S A 102: 18414–18419.
48. Chaves M, Albert R, Sontag ED (2005) Robustness and fragility of Boolean models for genetic regulatory networks. J Theoretical Biology 235: 431–449.
49. Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, et al. (2006) A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell 10: 515–527.
50. Bray D (1995) Protein molecules as computational elements in living cells. Nature 376: 307–312.
51. Iyengar R, Birnbaumer L, editors (1989) G Proteins. New York: Academic Press.
52. Jordan JD, Landau EM, Iyengar R (2000) Signaling networks: the origins of cellular multitasking. Cell 103: 193–200.
53. Eungdamrong NJ, Iyengar R (2004) Modeling cell signaling networks. Biology Cell 96: 355–362.
54. Eungdamrong NJ, Iyengar R (2004) Computational Approaches for modeling regulatory cellular networks. Trends Cell Biology 14: 661–669.
55. Gianchandani EP, Brautigan DL, Papin JA (2006) Systems analyses characterize integrated functions of biochemical networks. Trends Biochemical Sci 31: 284–291.
56. Blinov ML, Faeder JR, Goldstein B, Hlavacek WS (2006) A network model of early events in epidermal growth factor receptor signaling that accounts for combinatorial complexity. BioSystems 83: 136–151.
57. Inoki K, Corradetti MN, Guan KL (2005) Dysregulation of the TSC-mTOR pathway in human disease. Nat Genet 37: 19–24.
58. Sarbassov DD, Ali SM, Sabatini DM (2005) Growing roles for the mTOR pathway. Curr Opin Cell Biol 17: 596–603.
59. Chen Y, Rodrik V, Foster DA (2005) Alternative phospholipase D/mTOR survival signal in human breast cancer cells. Oncogene 24: 672–679.
60. Makris C, Voisin L, Giasson E, Tudan C, Kaplan DR, et al. (2002) The Rb-family protein p107 inhibits translation by a PDK1-dependent mechanism. Oncogene 21: 7891–7896.

The FASEB Journal article fj.07-9100com. Published online January 2, 2008.

Network topology determines dynamics of the mammalian MAPK1,2 signaling network: bifan motif regulation of C-Raf and B-Raf isoforms by FGFR and MC1R

Melissa Muller, Mandri Obeyesekere, Gordon B. Mills, and Prahlad T. Ram

Department of Systems Biology, University of Texas M.D. Anderson Cancer Center, Houston, Texas, USA

ABSTRACT **Activation of the fibroblast growth factor (FGFR) and melanocyte stimulating hormone (MC1R) receptors stimulates B-Raf and C-Raf isoforms that regulate the dynamics of MAPK1,2 signaling. Network topology motifs in mammalian cells include feed-forward and feedback loops and bifans where signals from two upstream molecules integrate to modulate the activity of two downstream molecules. We computationally modeled and experimentally tested signal processing in the FGFR/MC1R/B-Raf/C-Raf/MAPK1,2 network in human melanoma cells; identifying 7 regulatory loops and a bifan motif. Signaling from FGFR leads to sustained activation of MAPK1,2, whereas signaling from MC1R results in transient activation of MAPK1,2. The dynamics of MAPK activation depends critically on the expression level and connectivity to C-Raf, which is critical for a sustained MAPK1,2 response. A partially incoherent bifan motif with a feedback loop acts as a logic gate to integrate signals and regulate duration of activation of the MAPK signaling cascade. Further reducing a 106-node ordinary differential equations network encompassing the complete network to a 6-node network encompassing rate-limiting processes sustains the feedback loops and the bifan, providing sufficient information to predict biological responses.—Muller, M., Obeyesekere, M., Mills, G. B., Ram., P. T. Network topology determines dynamics of the mammalian MAPK1,2 signaling network: bifan motif regulation of C-Raf and B-Raf isoforms by FGFR and MC1R. *FASEB J.* 22, 000–000 (2008)**

Key Words: signaling motifs · computational modeling · proteomics · melanoma

B-Raf is mutationally activated in 60–80% of malignant melanomas as well as a large number of benign nevi, indicating a role in the initiation of malignant melanoma. Melanoma is the most aggressive form of skin cancer. Recently, protein kinase inhibitors have demonstrated remarkable clinical benefit in diseases that have been resistant to traditional chemotherapy, including chronic myelogenous leukemia (CML), gastrointestinal stromal tumors (GISTs), *HER2/Neu*-amplified breast cancer, and renal cell carcinoma (1–8). Each of these diseases is characterized by genetic aberrations that activate protein signaling networks, which appears to be critical to the efficacy of protein kinase inhibitors. Activating mutations of B-Raf result in constitutive activation (phosphorylation) of MAPK (9, 10). This pathway is also activated by mutation of N-Ras, which occurs in ~15% of melanomas (11, 12). Mutant N-Ras activates the RAS-RAF-MEK-MAPK as well as the PI3K-AKT-mTOR pathways *in vitro*. Mutations of B-Raf and N-Ras appear to be mutually exclusive in melanoma tumors and cell lines (12–14).

The high prevalence of activating mutations of components of the RAS-RAF-MEK-MAPK pathway suggests that it may be an effective therapeutic target in melanoma. The first B-Raf inhibitor to be used in clinical trials is sorafenib (Nexavar®), also known as BAY43–9006 (15). Sorafenib is a small molecule inhibitor of wild-type B-Raf, mutant (V600E) B-Raf, and a number of tyrosine kinase receptors (16). A Phase II single-agent study in patients with metastatic melanoma yielded disappointing results (16). Among 20 patients, only 1 partial response was observed, and 3 patients had stable disease. The lack of a complete understanding of the underlying homeostatic mechanisms regulating the RAS/RAF/MEK/MAPK pathway and the effects of B-Raf mutations on this pathway may contribute to the failure of monotherapy targeting individual components of the pathway. We thus sought to further define the homeostatic mechanisms controlling information transfer through the RAS/RAF/MEK/MAPK pathway (17).

Analysis of mammalian signaling networks shows that a large percentage of signaling subnetwork motifs are feed-forward/feedback loops and bifans (18). A recent analysis of network motifs of a 545 component 1259 interaction mammalian signaling network revealed 300 feed-forward loops and 1000 bifan subnetworks (18). The role these network motifs play in regulating signal-

[1] Correspondence: Department of Systems Biology, Unit 950, UT M.D. Anderson Cancer Center, 7435 Fannin St., Houston, TX 77054, USA. E-mail: pram@mdanderson.org

doi: 10.1096/fj.07-9100com

ing is not entirely clear. These subnetwork motifs could serve both to prolong signaling on initiation by a stimuli and also to terminate the signal (18–20). Given the high frequency of occurrence of bifan motifs in signaling and transcriptional networks (18, 21), we investigated whether coherent or incoherent bifan motifs regulate the mammalian MAPK1,2 network.

Signaling networks receive inputs from multiple upstream stimuli or receptor and must appropriately process this information into a clear message that results in the appropriate cellular outcomes. The MAPK1,2 network can be activated by signals from RTKs that activate Ras–C-Raf (22) and also by GPCR signals that activate B-Raf in a cAMP-dependent manner (23, 24) (see Supplemental Figs. S1 and S2). Signals from the Gαs–cAMP pathway also activate protein kinase A (PKA), which in turn phosphorylates C-Raf at Ser-259 and causes C-Raf to bind 14–3-3 proteins rendering C-Raf inactive (25–28). Identified feedback loops in this network include a positive feedback loop from MAPK1,2 involving PKC and potentially C-Raf (20), a negative loop from MAPK1,2 to B-Raf (29), a negative feedback loop from MAPK to C-Raf (30), a positive feedback loop from MAPK to C-Raf (31), and a feed-forward loop from PKC to adenylyl cyclase (32) (See Supplemental Figs. S2*A*, *B* and S3*A*, *B*, *D* and Fig. 2*A*). In summary, these interactions form a partially incoherent bifan, where C-Raf receives a positive and a negative input from FGFR and MC1R, respectively, while B-Raf receives positive inputs from both receptors. Given this complexity in connectivity within the GPCR-RTK-MAPK network, it is *a priori* difficult to understand how signals are processed to modulate the temporal duration of MAPK1,2 activity and what molecules and motifs can regulate the response. However, a computational model incorporating these concepts offers the potential to both characterize the dynamics of signal transduction through a network and define combinational therapies that could improve patient outcome.

Sustained *vs.* transient activation of signaling molecules and pathways is an important regulatory mechanism within the cell. For example, sustained activation of MAPK1,2 (90 min or more) has been shown to increase gene expression and activation of c-Fos leading to cellular proliferation, whereas the transient activation of MAPK1,2 (less than 30 min) does not activate c-Fos (33, 34). Similar effects are seen in NFkB signaling, where a transient or a sustained activation results in distinct changes in gene expression (35). Physiological responses such as differentiation and cell division are often dependent on temporal duration of activation of signaling molecules such as MAPK1,2 (36, 37). We have previously shown that the MAPK1,2 network can exists as a bistable system and can elicit either a sustained or transient response (20).

In this study, we determined how bifan and loop motifs regulate the system properties of the mammalian MAPK1,2 network in human melanoma cells. To address this problem, we used an integrated approach of computational modeling and experimental analysis. For the experimental analysis, we used SB2 human melanoma cells. This model system and cell line were chosen for several reasons. Melanocytes and melanoma cell lines endogenously express both fibroblast growth factor receptor (FGFR) (38) as well melanocortin 1 receptor (MC1R) (39), which are important in their biology and the pathophysiology of melanoma. Fibroblast growth factor is an important mitogen for melanoma cells (40), and the melanocortin 1 receptor, which is coupled to Gs, is important in normal biology as well as in the pathophysiology of melanoma (41). Single-nucleotide polymorphisms of the MC1R that result in a loss-of-function variant of the receptor are a risk factor for the development of melanoma (42). At the signaling level, B-Raf mutations occur in ~70% of melanomas, and B-Raf is a target for therapy (10). The cell line we have used in this study endogenously expresses FGFR and MC1R, as well as wild-type B-Raf and C-Raf, and contains the V90 MC1R loss-of-function SNP. The computational model encompassed ordinary differential equations (ODE) to simulate the MAPK1,2 signaling network.

MATERIALS AND METHODS

Computational modeling

Two computational models were developed: Model A, a detailed 106 ODE system; and Model B, a reduced 6 ODE system. A previously published MAPK1,2 network model (20) and a published Gs-PKA module (43) were integrated to form Model A. The same parameter values and rate constants were used, except for FGF and FGFR, the values for which are in Table S1. These two models were connected by the following interactions: *1*) cAMP binding and activation of cAMP-GEF; *2*) cAMP-GEF activation of Ras; *3*) Ras activation of B-Raf; *4*) B-Raf activation of MEK1; *5*) PKA phosphorylation of Ser-259 of C-Raf; and *6*) PP2A dephosphorylation of Ser-259 C-Raf. The connection from cAMP-GEF to B-Raf is mediated by a small G protein. The small G protein it is most commonly reported as Rap in many cell types. However, in melanoma it is not clear whether cAMP-GEF activates Rap or Ras. Regarding melanoma cells, Busca *et al.* (24) and Dumaz *et al.* (9) present data to suggest that Ras mediates the cAMP signal; whereas Gao *et al.* (44) suggest that Rap mediates the cAMP signal. Since the connection from cAMP-GEF to Rap/Ras is not well defined in melanoma, we chose to use the cAMP-GEF to Ras connection for the model. A coupled ordinary differential equation computational model of the network was developed and solved in Genesis/Kinetikit (19). Parameter variation analysis was done on the network through iterations of modeling and experimental analysis. A parameter set was chosen such that the model output closely matched the experimental data. These parameter values were used for subsequent simulations. The biochemical parameters and concentrations that were used for the added components are shown in Supplemental Table S1. The second model, Model B, of the reduced network was built and solved using MatLab. The parameters and constants are shown in Supplemental Table S2.

Cell culture and stimulation

Human SB2 melanoma cells were kindly provided by Dr. Menashe Bar-Eli (M.D. Anderson Cancer Center, Houston, TX, USA). SB2 melanoma cells endogenously express wild-type B-Raf, C-Raf, FGFR, and the V90 MC1R SNP (unpublished results, Drs. J. Ellerhorst and E. Grimm, Department of Experimental Therapeutics, University of Texas M.D. Anderson Cancer Center, Houston, TX, USA). Cells were routinely maintained in modified essential medium supplemented with 10% FBS. For signaling experiments, logarithmically growing cells were starved in 1% FBS for 16 h and then subjected to stimulations using basic fibroblast growth factor (FGF 20 ng/ml) (Cell Signaling Technology, Danvers, MA, USA) and/or melanocyte stimulating hormone (MSH, 2 μM; Sigma-Aldrich, St. Louis, MO, USA). For washout experiments, cells were stimulated with fibroblast growth factor (FGF) and/or MSH for 5 min, washed, and incubated at 37°C in 1% medium for time periods indicated in figure legends. Where specified, cells were pretreated with the PKA inhibitor H89 (20 μM) for 2 h prior to incubation with FGF and/or MSH. Controls were incubated for corresponding times with dimethyl sulfoxide. For some experiments, cells were treated with short interfering RNA (siRNA) for 48 h prior to treatment to knock down C-Raf (Cell Signaling Technology).

Antibodies

The following antibodies were used for immunoblotting and reverse phase protein microarrays (RPPAs): antiphospho-p44/42 MAPK, p42MAPK (Cell Signaling Technology); anti-C-Raf (Upstate Biotechnology, Waltham, MA, USA); and anti-β-Actin (Sigma-Aldrich).

SDS-PAGE and immunoblotting

Cells were lysed by incubation on ice for 15 min in a sample lysis buffer (50 mM HEPES; 150 mM NaCl; 1 mM EGTA; 10 mM sodium pyrophosphate, pH 7.4; 100 nM NaF; 1.5 mM $MgCl_2$; 10% glycerol; and 1% Triton X-100 plus protease inhibitors aprotinin, bestatin, leupeptin, E-64, and pepstatin A). Cell lysates were centrifuged at 15,000 *g* for 20 min at 4°C. The supernatant was frozen and stored at −20°C. Protein concentrations were determined using a protein-assay system (BCA, Bio-Rad, Hercules, CA, USA), with BSA as a standard. For immunoblotting, proteins (25 μg) were separated by SDS-PAGE and transferred to Hybond-C membrane (GE Healthcare, Piscataway, NJ, USA). Blots were blocked for 60 min and incubated with primary antibodies overnight, followed by goat anti-mouse IgG-HRP (1:30,000; Cell Signaling Technology) or goat anti-rabbit IgG-HRP (1:10,000; Cell Signaling Technology) for 1 h. Secondary antibodies were detected by enhanced chemiluminescence (ECL) reagent (GE Healthcare). All experiments were repeated a minimum of three independent times.

RPPA

Lysates were prepared as for Western blotting. Cell lysates (1 μg/μl) were boiled in 1% SDS and hybridized under stringent conditions mimicking the conditions used in Western blotting. Using a GeneTac G3 DNA arrayer (Genomic Solutions, Ann Arbor, MI, USA), seven two-fold serial dilutions of cell lysates are arrayed on multiple nitrocellulose-coated glass slides (FAST Slides, Whatman Schleicher & Schuell, Keene, NH, USA). RPPA slides were produced in batches of 20. Printed slides were stored in desiccant at −20°C. Antibodies were screened for specificity by Western blotting with 25 μg of lysate protein per lane. An antibody was accepted only if it produced a single predominant band at the expected molecular weight and where proteins behaved similarly on Western blotting and arrays following manipulation of signaling pathways or across multiple cell lines with a wide dynamic range. Each array was incubated with specific primary antibody, which was detected by using the catalyzed signal amplification (CSA) system (DAKO, Carpinteria, CA, USA). Briefly, each slide was washed in a mild stripping solution of Re-Blot Plus (Chemicon International, Temecula, CA, USA) then blocked with I-block (Tropix, Bedford, MA, USA) for at 30 min. Following the DAKO universal staining system, slides were then incubated with hydrogen peroxide, followed by avidin for 5 min, and biotin for 5 min. Slides were incubated with primary and secondary antibodies then incubated with streptavidin-peroxidase for 15 min, biotinyl tyramide (for amplification) for 15 min, and 3,3-diaminobenzidine tetrahydrochloride chromogen for 5 min. Between steps, the slide was washed with TBS-T buffer. Loading is determined by comparing phosphorylated and nonphosphorylated antibodies. Multiple controls are placed on each slide to facilitate quantification and validation of the assay. Spot intensity was measured using Image J (http://rsb.info.nih.gov/ij/l). Protein phosphorylation levels are expressed as a ratio to equivalent total proteins. Fold increases in spot intensities were calculated against nonstimulated control samples.

RESULTS

MSH transiently activates MAPK1,2

A detailed 106 ODE model of the MAPK1,2 network receiving inputs from FGF and MSH through the FGFR and MC1R was constructed (Model A). Model A includes two previously published subnetworks, the RTK-MAPK1,2 (20) and the Gs-PKA (43) subnetworks connected through cAMP-GEF activation of Ras-B-Raf and PKA inhibition of C-Raf. The table of molecules, initial concentrations, interactions, and constants are shown in Supplemental Table S1. The network includes a putative bifan, two positive feedback loops, three negative feedback loops, and two feed-forward loops based on published literature (18, 21, 45). We had previously shown that activation with platelet-derived growth factor (PDGF) leads to a sustained activation of MAPK1,2 (20). Initially, investigations were performed to determine the duration of MAPK1,2 activity in response to a stimulus received through activation of a GPCR (MC1R). Computational modeling of the MAPK1,2 network with Model A, predicted that a 5-min stimulus of 0.02 μM MSH would lead to only a transient increase in MAPK1,2 phosphorylation with a return to near basal levels 10 min after stimulus (**Fig. 1*A***) if a positive feedback loop exists from MAPK to C-Raf. The MSH stimulus was predicted to result in a maximal activation of 0.16 μM of active MAPK1,2, 5 min after stimulus. We have operationally defined a sustained activation as being at least 6 times longer than the stimuli. In our computational and experimental studies, the length of stimuli was 5 min and simulations were made up to 60 min after stimuli. However, computational modeling of a positive feedback loop to both C-Raf and B-Raf from

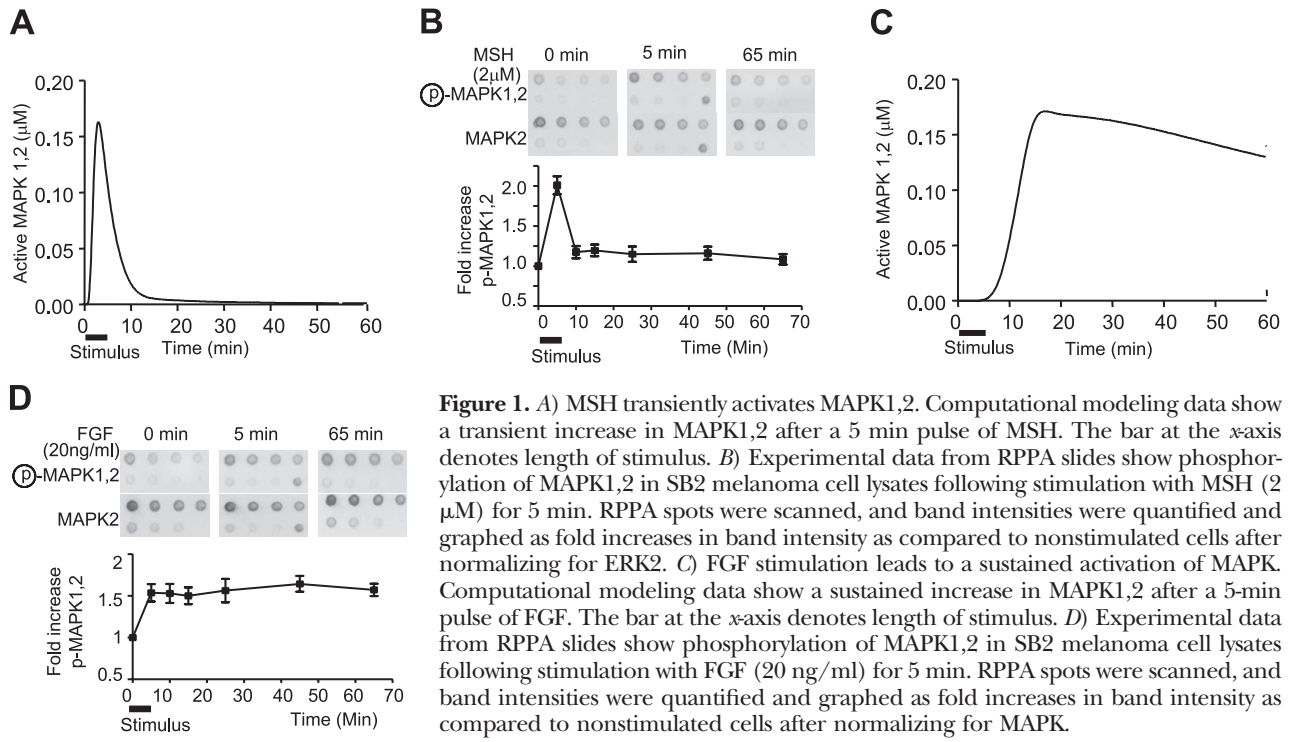


Figure 1. *A*) MSH transiently activates MAPK1,2. Computational modeling data show a transient increase in MAPK1,2 after a 5 min pulse of MSH. The bar at the *x*-axis denotes length of stimulus. *B*) Experimental data from RPPA slides show phosphorylation of MAPK1,2 in SB2 melanoma cell lysates following stimulation with MSH (2 μM) for 5 min. RPPA spots were scanned, and band intensities were quantified and graphed as fold increases in band intensity as compared to nonstimulated cells after normalizing for ERK2. *C*) FGF stimulation leads to a sustained activation of MAPK. Computational modeling data show a sustained increase in MAPK1,2 after a 5-min pulse of FGF. The bar at the *x*-axis denotes length of stimulus. *D*) Experimental data from RPPA slides show phosphorylation of MAPK1,2 in SB2 melanoma cell lysates following stimulation with FGF (20 ng/ml) for 5 min. RPPA spots were scanned, and band intensities were quantified and graphed as fold increases in band intensity as compared to nonstimulated cells after normalizing for MAPK.

MAPK1,2 shows a biphasic activation of MAPK1,2 in response to MSH (Supplemental Fig. S1*C*). Given these two different potential outputs from Model A, we experimentally determined the MAPK1,2 response to MSH stimulation in SB2 melanoma cells. SB2 melanoma cells were serum-starved for 16 h and stimulated with 2 μM MSH for 5 min. The cells were washed with serum-free medium and incubated for various lengths of time. The resulting phosphorylation of MAPK1,2 was determined by RPPA analysis. Experimental data show that MSH stimulation results in peak activity of phosphorylated MAPK1,2, 5 min after stimulus with a return to basal levels 10 min after stimulus (Fig. 1*B*). Similar results were observed in MeWo human melanoma cells (data not shown). The same sets of lysates were probed for phospho-MAPK1,2 and total MAPK on a Western blot confirming that MAPK1,2 phosphorylation is only transiently increased in response to MSH (Supplemental Fig. S3*A*). Therefore, based on the experimental results we eliminated the positive feedback from MAPK1,2 to B-Raf from the model and operated with a single positive feedback loop to C-Raf. The iterative use of computational simulation and experimental data is thus used to test and constrain the model.

FGF leads to a sustained activation of MAPK1,2

We have previously shown that stimulation by PDGF leads to a sustained activation of MAPK1,2 in mouse fibroblast cells (20). We wanted to determine whether other growth factors that activate receptor tyrosine kinases, such as bFGF, can also elicit a sustained activation of MAPK1,2 in melanoma cells. Computational modeling of the network using the revised Model A, with a positive feedback from MAPK1,2 to C-Raf, suggested that a brief 5-min pulse of FGF would lead to a sustained increase in phosphorylated MAPK1,2 (Fig. 1*C*) with a maximal increase of activated MAPK1,2 of 0.17 μM. Additional reported pathway connection within the network include a negative feedback from MAPK1,2 to PKC (46). We modeled these interactions individually and simulated the MAPK1,2 response with these different connections. The results from these simulations show that including a negative feedback from MAPK to PKC results in a more rapid reduction in MAPK1,2 activity as compared to the simulations with only a positive feedback loop (Supplemental Fig. S2*C*). Interestingly, adding a feed-forward loop from PKC to AC (Supplemental Fig. S2*D*) changed the predicted MAPK1,2 activation to a biphasic response (Supplemental Fig. S2*E*). Given these different computational predictions, we determined the effects of bFGF on SB2 melanoma cells. Cells were serum-starved and stimulated with bFGF (20 ng/ml) for 5 min. The FGF was washed out, and cells were incubated for different lengths of time. The RPPA data show that a brief stimulus of FGF leads to a sustained increase in MAPK1,2 phosphorylation (Fig. 1*D*) with a maximal increase in phosphorylated MAPK1,2, 10 min after stimulus. Similar results were obtained using Western blot analysis (Supplemental Fig. S3*B*). Constraining the model based on the observed biological data, we eliminated the feed-forward loop from PKC to AC (Supplemental Fig. S2*D*), as well as the negative feedback loop from MAPK1,2 to C-Raf (Supplemental Fig. S2*B*). The resulting Model A used for further work contains a positive feedback loop from MAPK1,2 to C-Raf (**Fig. 2*A***).

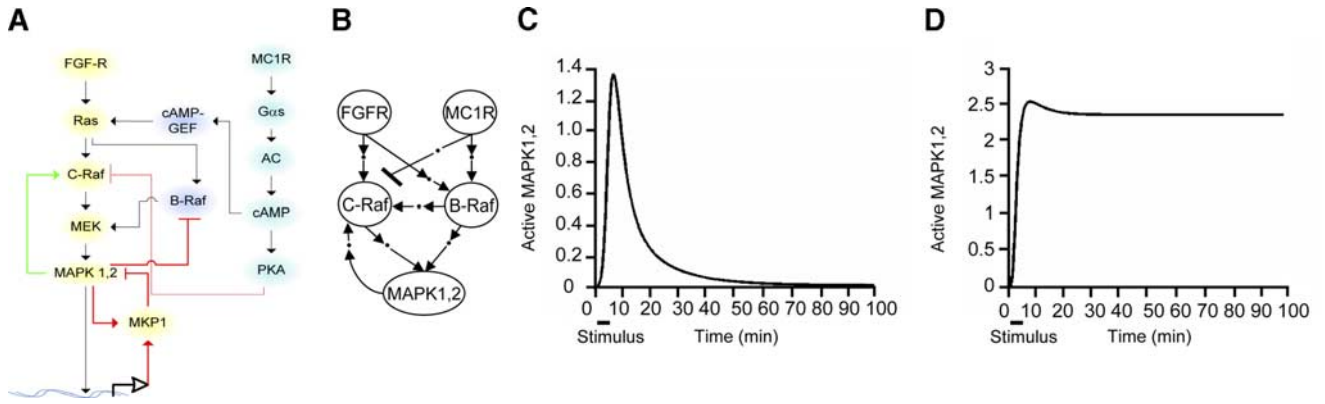


Figure 2. *A*) Detailed network model of the FGFR-MC1R-B-Raf-C-Raf-MAPK1,2 network based on the experimental data from Fig. 1. *B*) Connection diagram shows the reduced network structure of the FGFR-MC1R-B-Raf-C-Raf-MAPK1,2 network dynamics of MAPK1,2 activity. *C*) The reduced motif model was trained using the experimental data from MSH stimulation. The model shows only a transient increase in active MAPK on a 5-min MSH pulse. *D*) The same model was also trained using the FGF experimental data. The model shows a sustained activation of MAPK on FGF stimulation.

The MC1R, FGFR, B-Raf, C-Raf bifan network

Based on the data from Fig. 1, the detailed network we developed is shown in Fig. 2*A*. This final network Model A (Fig. 2*A*) was used for the remainder of the work in this paper. The various interactions within the MAPK1,2 network predicts that signaling from MC1R and FGFR to B-Raf and C-Raf forms a partially incoherent bifan motif. This bifan motif is described as partially incoherent since B-Raf has coherent inputs (both positive from MC1R and FGFR), while C-Raf is incoherent (positive from FGFR and negative from MC1R). The bifan drives a positive feedback loop from MAPK,1,2 to C-Raf, this network structure is shown in Fig. 2*B*.

Reduction of network and computing the signal processing based on network structure

Based on connectivity to B-Raf and C-Raf we wanted to determine whether a reduced network could be developed into a predictive computational model. This reduction is a very important issue in the development of large quantitative predictive models of signaling networks. Our detailed ODE model already contains 106 equations, and adding on any other signaling pathways would very rapidly increase the model to several hundred ODEs. Computing hundreds of ODE's increases the likelihood of errors and, with the difference in time scales, can lead to stiff systems. We thus determine whether reducing the model to include only nodes that integrate signals would decrease the number of equations required to solve, while still maintaining the ability to predict experimental outcomes. Preserving the signatures of the network, feedbacks, and bifans, the network was reduced to the minimum informative network structure. A reduced network model was developed, Model B (Supplemental Fig. S5), and tuned using the experimental data from Fig. 1*B*, *D*. The simulation from the reduced model predicts a transient activation of MAPK1, 2 on MSH stimulus (Fig. 2*C*). Model B was constrained using the experimental data with resulting simulations predicting that activation of FGFR results in sustained activation of MAPK1,2 (Fig. 2*D*). Having developed this reduced model and tuning it using the experimental data, we used Model B to predict the behavior of the system and compared it to predictions from the large, detailed Model A in the ensuing sets of studies.

C-Raf is essential for a sustained but not transient activation of MAPK1,2

Regulation of the temporal characteristics of the MAPK1,2 network by the two RAF isoforms is not known. Reports suggest that C-Raf is needed for activation of MAPK1,2 by B-Raf stimulatory pathways (47, 48). We wanted to investigate how changing the relative protein levels of C-Raf and B-Raf would alter the MAPK1,2 response. Parameter variation of the concentrations of B-Raf and C-Raf was performed with the resultant computational data suggesting that the MAPK1,2 response is acutely sensitive to the relative concentrations of the two Raf isoforms (**Fig. 3*A***). Sensitivity analysis suggests that the relative amount of C-Raf would be critical for eliciting a sustained MAPK1,2 response. If the initial C-Raf protein concentration is equal to or higher than the B-Raf protein concentration, the MAPK1,2 network is predicted to respond to stimulus resulting in a sustained activation of MAPK1,2 (Fig. 3*A*). We experimentally determined the relative expression of C-Raf and B-Raf in SB2 melanoma cells. Immunoprecipitation of C- and B-Raf followed by Coomassie blue staining of the immunoprecipitate showed that the expression levels of C-Raf and B-Raf are almost equal (Fig. 3*B*). Closer examination of the MAPK1,2 response suggests that a steep C-Raf concentration dependency would be required to elicit a sustained MAPK1,2 response (Fig. 3*C*). Based on Model B, we simulated the FGF activation of MAPK

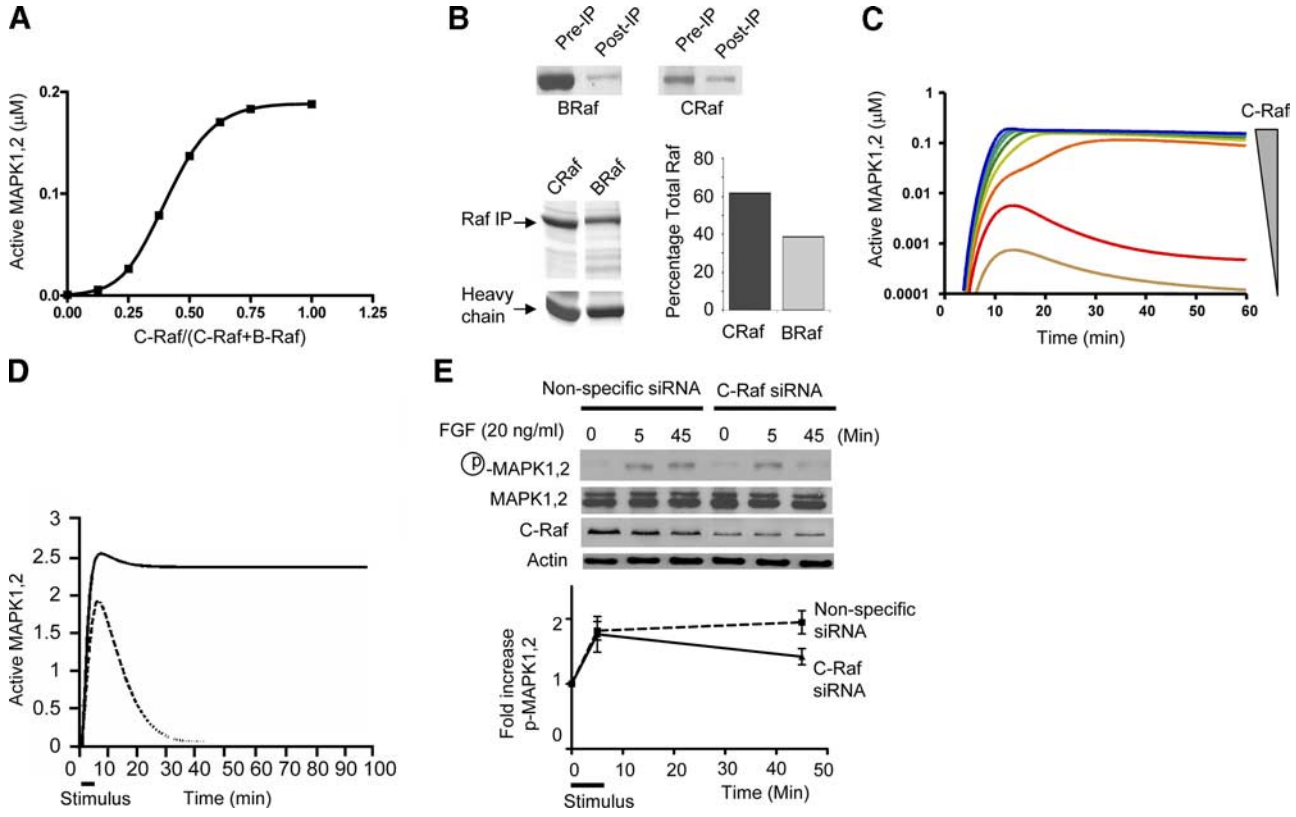


Figure 3. *A*) C-Raf is essential for sustained MAPK1,2 activation. Computational modeling of FGF activation of MAPK1,2 activity 30 min after stimulus with different ratios of C-Raf to B-Raf. *B*) Experimental data show immunoprecipitated B-Raf and C-Raf from SB2 melanoma cells resolved on SDS-PAGE and stained with Coomassie blue. Band intensities were quantified and shown on the bar graph as relative expression of B-Raf and C-Raf after normalizing against the IgG heavy chain and for the change in B-Raf and C-Raf after immunoprecipitation, as compared to the levels in the pre-IP lysate. *C*) Sensitivity analysis of the sustained *vs.* transient response of MAPK1,2 activity to FGF as a function of varying the total amount of C-Raf suggests that the sustained MAPK1,2 response is dependent on C-Raf levels. *D*) The reduced motif model shows inhibition of the activation of C-Raf by MAPK (dotted line) when the biological system is stimulated by FGF for 5 min. *E*) Experimental data from SB2 cells transfected with either C-Raf specific siRNA or nonspecific siRNA. Seventy-two hours after transfection, cells were serum-starved and stimulated with FGF for 5 min, washed out, and incubated for the times shown. Cell lysates were probed for phospho-MAPK1,2, C-Raf, and β-Actin as shown. Graph shows changes in phopsho-MAPK1,2 after normalization for β-Actin.

under conditions where the feedback to C-Raf was inhibited. The modeling simulation predicted only a transient activation of MAPK when feedback to C-Raf is blocked (Fig. 3*D*). These computational modeling results predict that if C-Raf levels are decreased, FGF will no longer elicit a sustained MAPK1,2 response (Fig. 3*D*; dotted line). To experimentally test this hypothesis, cells were transfected with C-Raf specific siRNA and the levels of C-Raf were measured 72 h after transfection. The data show that C-Raf was knocked down by 60% using C-Raf siRNA with little effect on B-RAF (Supplemental Fig. S3*D*). We then determined the MAPK1,2 response to stimulation by FGF under conditions of low C-Raf. Cells transfected with either C-Raf or nonspecific siRNA were serum-starved and stimulated with FGF for 5 min and the phosphorylation of MAPK1,2 measured over time. The data show that if C-Raf is knocked down, FGF can no longer elicit a sustained MAPK1,2 response, while a transient response is still present (Fig. 3*E*). These data suggest that the expression level of C-Raf is critical for a sustained MAPK1,2 response.

The bifan motif can regulate the duration of MAPK1,2 activation

Having observed the differential regulation of the dynamics of MAPK1,2 activity by C-Raf, we wanted to determine whether the duration of MAPK1,2 activation can be modulated in a completely endogenous system. Activation of the Gs-PKA pathway phosphorylates C-Raf and inhibits C-Raf activity (28). Therefore, we wanted to determine whether activation of the Gs-PKA pathway would gate the FGF-induced sustained activation of MAPK1,2 by inhibiting C-Raf. Computational modeling data from Model A (Fig. 2*A*) predicted that a prepulse of MSH would inhibit FGF-induced sustained phosphorylation of MAPK1,2 (**Fig. 4*A***). While a pulse of FGF alone activated MAPK1,2 for longer than 70 min (Fig. 1*C*), a prepulse of MSH followed by a pulse of FGF phosphorylated MAPK1,2 for less than 30 min (Fig. 4*A*). Model B also predicted a transient activation of MAPK on a prepulse of MSH followed by FGF (Fig. 4*B*). SB2

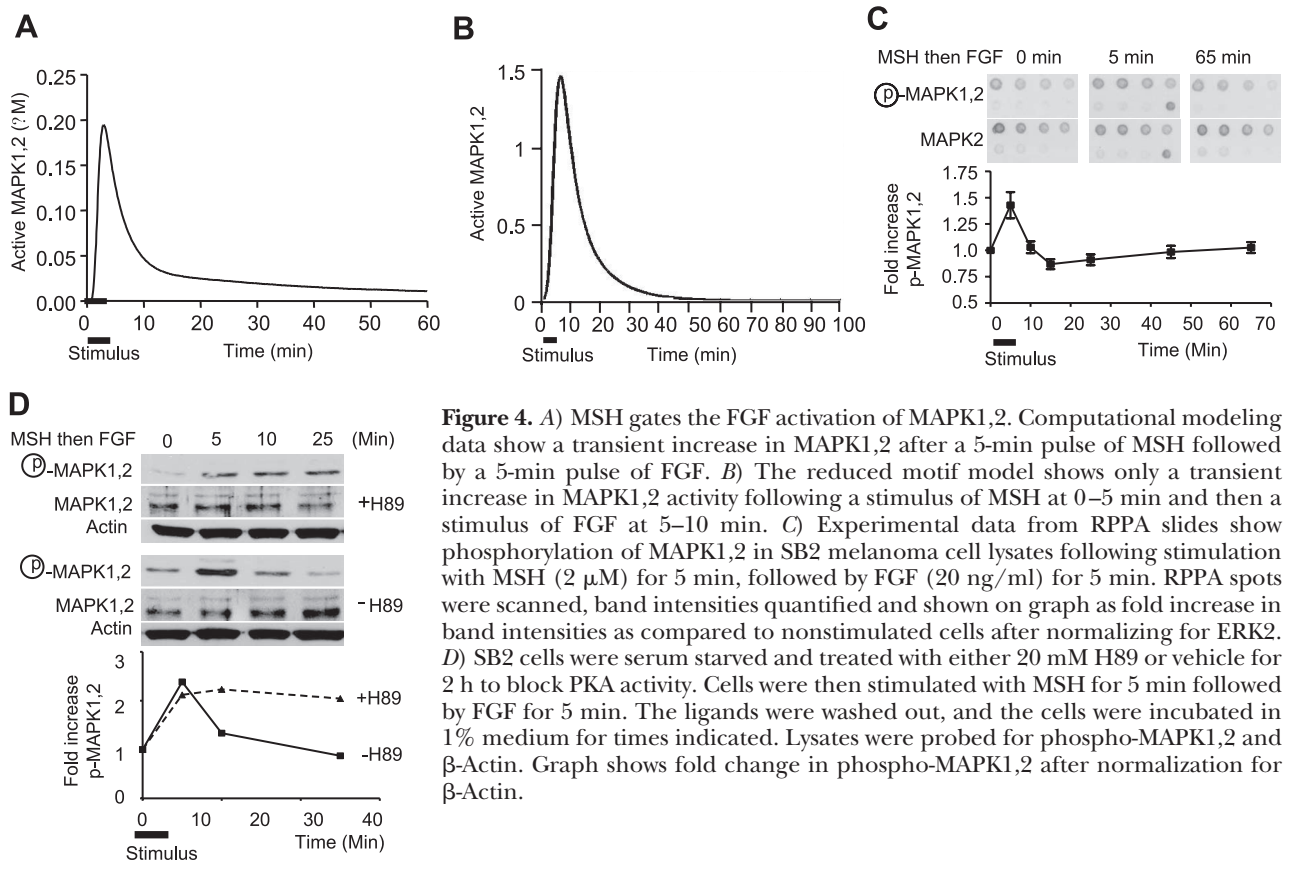


Figure 4. *A*) MSH gates the FGF activation of MAPK1,2. Computational modeling data show a transient increase in MAPK1,2 after a 5-min pulse of MSH followed by a 5-min pulse of FGF. *B*) The reduced motif model shows only a transient increase in MAPK1,2 activity following a stimulus of MSH at 0–5 min and then a stimulus of FGF at 5–10 min. *C*) Experimental data from RPPA slides show phosphorylation of MAPK1,2 in SB2 melanoma cell lysates following stimulation with MSH (2 μM) for 5 min, followed by FGF (20 ng/ml) for 5 min. RPPA spots were scanned, band intensities quantified and shown on graph as fold increase in band intensities as compared to nonstimulated cells after normalizing for ERK2. *D*) SB2 cells were serum starved and treated with either 20 mM H89 or vehicle for 2 h to block PKA activity. Cells were then stimulated with MSH for 5 min followed by FGF for 5 min. The ligands were washed out, and the cells were incubated in 1% medium for times indicated. Lysates were probed for phospho-MAPK1,2 and β-Actin. Graph shows fold change in phospho-MAPK1,2 after normalization for β-Actin.

melanoma cells were serum-starved and stimulated with 2 μM MSH for 5 min, followed by 20 ng/ml FGF for 5 min. Ligands were then washed out, and cells were incubated in serum-free media for different lengths of time. The data show that a prepulse of MSH does indeed inhibit the FGF-induced sustained activation of MAPK1,2 (Fig. 4*C*). Similar data were obtained from Western blot analysis (Supplemental Fig. S3*C*).

PKA is a key regulator determining the state of the MAPK1,2 network

The data thus far show that the levels and activation of specific Raf isoforms determine the temporal duration of activation of the MAPK1,2 network, whereby either knocking down C-Raf or inhibiting its function by activating PKA inhibits a sustained MAPK1,2 response. We further investigated the role of PKA in the MSH gating of the FGF-mediated MAPK1,2 network. The MSH regulation of MAPK1,2 is dependent on cAMP and PKA, where cAMP has been reported to activate B-Raf (24, 49) but PKA inhibits C-Raf (50). This divergence of the signal occurs on two different hierarchical steps, where cAMP is at a level higher than PKA. Therefore, it should be possible to inhibit PKA activity using a pharmacological inhibitor (H89), while at the same time allowing cAMP activation of B-Raf. Under this condition, we determined whether inhibiting PKA, and therefore not inhibiting C-Raf, would alter the MSH gating of MAPK1,2. SB2 melanoma cells were serum-starved, and PKA activity was blocked with the incubation of H89. Cells were stimulated with MSH for 5 min followed by FGF for 5 min, and the activity of MAPK1,2 was measured. The data show that if PKA is inhibited, MSH cannot gate the FGF activation of MAPK1,2 (Fig. 4*D*).

Inhibiting the negative input into the bifan switches MAPK1,2 activation from transient to sustained

Computational modeling using Model A (Fig. 2*A*), of the MSH activation of MAPK1,2 under conditions where PKA is inhibited predicted that MSH stimulation would lead to a sustained activation of MAPK1,2 when PKA activity is blocked (**Fig. 5*A***). SB2 melanoma cells were serum-starved and incubated with 20 μM H89 for 2 h to inhibit PKA activity. The cells were then stimulated with MSH for 5 min, and MAPK1,2 activity was measured over time. The data show that inhibiting PKA with H89 changes the MSH activation of MAPK1,2 from a transient response (Fig. 1*B*) to a sustained activation of MAPK1,2 in melanoma cells (Fig. 5*B*).

DISCUSSION

Signaling networks are complex with numerous subnetwork motifs that can dramatically alter the activation of important regulatory molecules. We constructed a

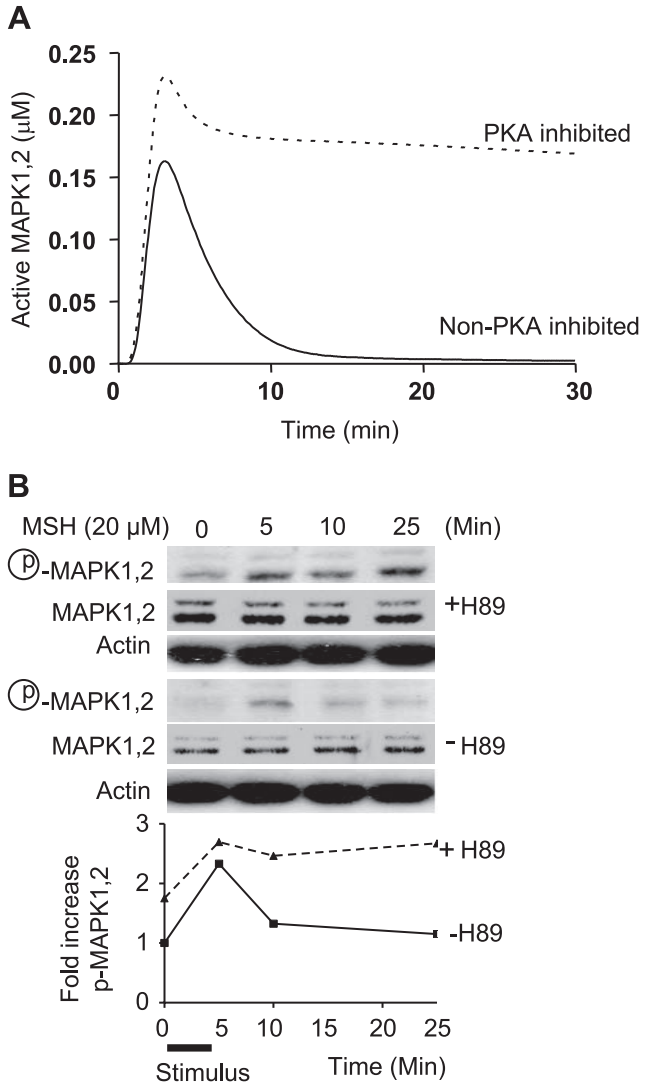


Figure 5. *A*) Computational modeling of the MSH activation of MAPK1,2 under control conditions (solid line) or with PKA activity inhibited (dashed line). *B*) SB2 cells were serum-starved and treated with either 20 mM H89 or vehicle for 2 h to block PKA activity. Cells were then stimulated with MSH for 5 min and washed out, and cells were incubated in serum-free medium for times indicated. Lysates were probed for phospho-MAPK1,2 and β-Actin. The graph shows fold change in phospho-MAPK1,2 after normalization for β-Actin.

MAPK1,2 network *in silico* based on our previous studies and on the literature and experimentally tested it to understand the design principles of this signaling module. The model was constrained based on experimental data to result in Model A (Fig. 2*A*). From these studies we show that the temporal duration of MAPK1,2 is distinct, depending on the activation signal. MC1R stimulation by MSH leads to only a transient activation of MAPK1,2. However, activation of FGFR leads to a sustained activation of MAPK1,2. These two incoming signals can interact such that the MC1R signal gates the FGFR signal to change it from a sustained to a transient activation of MAPK1,2 when the signals are processed through the partially incoherent bifan motif. The FGFR signal is coherent within the bifan to activate both C-Raf and B-Raf. C-Raf then couples to a positive feedback loop from C-Raf to MAPK1,2 back to C-Raf, resulting in a sustained activation of MAPK1,2. However, the MC1R signal is incoherent within the bifan where it activates B-Raf but inhibits C-Raf. This condition results in an inhibition of the C-Raf coupled feedback loop to and from MAPK1,2. The difference we observe in the sustained *vs.* transient activation of MAPK1,2 is explained by this difference in the connection within the bifan. While it is known that both B-Raf and C-Raf can activate MAPK1,2 (12, 13), here we show that the two isoforms have distinct roles in regulating the systems response of MAPK1,2. The system's response of MAPK1,2 activation is dependent on the nature of connectivity of the two receptors to B-Raf and C-Raf, which form a partially incoherent bifan. The incoherent bifan is the locus of signal integration from RTK and GPCR to MAPK1,2 with C-Raf acting as a logic gate. Computational modeling of the network of functional connections predicted the behavior of the system.

The bifan network controls the state of a positive feedback loop downstream from C-Raf to regulate the dynamics of MAPK1,2 activation. We have previously described the bistable properties of the MAPK1,2 network, which is dependent on a positive feedback loop from MAPK1,2 to C-Raf (20). Here we show that altering the connectivity to the bifan motif changes the state of MAPK1,2 from a transient to a sustained response. Removing the inhibitory or incoherent connectivity by inhibiting PKA switches the MC1R mediated transient activation of MAPK1,2 to a sustained response. Activation of PKA by GPCR's serves to limit the duration of MAPK1,2 activity, and ligands that activate Gs receptors can balance growth factor activation of MAPK1,2 (18). The FGFR signal leads to a sustained activation of MAPK1,2. However if C-Raf levels are knocked down or if C-Raf activity is inhibited, FGFR only elicits a transient activation of MAPK1,2. While FGFR can still transiently activate MAPK1,2 when C-Raf is knocked down, the sustained activation of MAPK1,2 is dependent on C-Raf. Our studies presented here show that sustained *vs.* transient activation of MAPK1,2 can be regulated by a feedback loop to C-Raf, which drives the systems response. PC12 rat pheochromocytoma cells also exhibit a temporal difference in MAPK phosphorylation in response to either EGF or NGF. Work by Santos *et al.* (51) using PC12 cells shows that a positive feedback loop from MAPK to C-Raf mediates sustained MAPK activity in response to NGF, while a negative feedback loop from MAPK to C-Raf leads to a transient activation of MAPK in response to EGF. Work by Sasagawa *et al.* (52), however, suggests that the differential regulation of either Ras or Rap1 by modulation of GTPase activating proteins (Ras-GAP or Rap-GAP) can lead to sustained *vs.* transient activation of MAPK in response to either EGF or NGF in PC12 cells. Here we show that a positive feedback to C-Raf and differential regulation of C-Raf *vs.* B-Raf in response to FGF or MSH can lead to transient *vs.* sus-

tained activation of MAPK in human melanoma cells. The two isoforms of Raf can activate MAPK1,2 with distinct temporal patterns of activity, suggesting that inhibiting C-Raf rather than B-Raf may have a greater effect on inhibiting MAPK1,2 activity. A previous modeling study of a linear pathway from B-Raf and C-Raf to MAPK1,2, which lacked the physiologically relevant feedback loops, suggested that a high concentration of B-Raf can lead to a sustained activation of MAPK1,2 (53) based on the parameters used in the model. We find that the concentration and activity of C-Raf is more important in maintaining the sustained response due to the feedback connectivity to C-Raf. Note that the immunoblot showed a darker band for B-Raf than C-Raf, but in fact when we immunodepleted the two Raf isoforms from aliquots of the same lysate and determined Raf levels by Coomassie blue staining, we observed that SB2 cells express more C-Raf than B-Raf. The role of C-Raf in maintaining the sustained MAPK activity may also explain why no C-Raf mutations have been found in melanoma. While B-Raf mutations are found in 70% of melanoma, B-Raf mutant expression by itself is not sufficient to drive transformation of cells but requires the addition of FGF (54). While both B-Raf and C-Raf can activate MAPK1,2 (12, 13), our studies here show for the first time that the two isoforms have distinct systems properties in differentially regulating the temporal duration of MAPK1,2 activation based on feedback loops and signaling motifs within the MAPK1,2 subnetwork. Activating mutations of B-Raf are highly prevalent in melanoma while activating mutations in C-Raf have not been reported. A possible reason for this is that while the B-Raf mutation increase MAPK1,2 by constant activation, the increase in MAPK1,2 phosphorylation is only ~2 fold (9), possibly keeping the cells in an increased proliferative state and thus increasing the chances of additional mutations. Note that B-Raf mutations are found in all stages, including preneoplastic nevi and melanocytic lesions, suggesting that B-Raf mutations may be an early event in the development of melanoma. Activating mutations of C-Raf have not been reported; one possible mechanistic reason could be that activating mutations of C-Raf may result in cell death by driving MAPK1,2 and the positive feedback loop.

Developing large quantitative models of signaling networks is very important as we try to predict how signals are processed in response to activators and change in response to targeted manipulations. Building detailed ODE models of large networks is impractical due to the scale of the model. Alternative methods such as Boolean models have been suggested, but due to feedback loops the Boolean models are frequently unstable and lead to oscillatory behavior. Here we demonstrate that reduction of the network based on the biology of the connectivity is a very useful method to reduce the network while maintaining all subtle complexities resulting in a model able to robustly predict the effects of network perturbations. Development of such systems biology methods is highly dependent on very close co-operation between mathematicians and biologists and experimental validation and constraining of the model is vital. Indeed, as indicated herein a number of different models were predicted from the literature. We tested these directly resulting in the evolution of Model A (Supplemental Fig. S1*A*) to the refined version of Model A depicted in Fig. 2*A*, which is considerably different in topology and in predictions.

Determining the systems function of signaling networks using an integrated approach of computational modeling, experimental biology, and targeted manipulations is very useful in identifying components and motifs that can regulate network function. These regulatory molecules could then serve as targets for therapeutic purposes. While a molecule might be an attractive target in the context of a linear pathway when one considers the feedback loops and interactions at a systems level, the same molecule may no longer provide to be a good target. Indeed, experimental and patient data for example suggest that a feedback loop in the phosphatidylinositol-3-kinase pathway limits the efficacy of targeting mTOR (55). Thus, understanding the systems properties and design principles of signaling networks and developing models of the relevant interactions validated with experimental testing could aid in development and validation of targeted therapeutics.

Medium-to-high throughput biological methods such as RPPA have allowed us to measure changes in several signaling molecules in parallel from the same set of cell lysates. Using these data, we can begin to construct larger networks and determine how signals from the cell surface are processed through the signaling network. Computational analysis has allowed us to reverse-engineer the system and direct some of the experimental work. Such analysis is important in allowing us to fully understand complex biological networks permitting reverse engineering to achieve desirable outputs from biological systems. FJ

The work was supported in part by startup funds from the M.D. Anderson Cancer Center, Melanoma SPORE P50 CA093459 (PP-CDP3), and U.S. Department of Defense grant BC044268 to P.T.R.

REFERENCES

1. Druker, B. J., Talpaz, M., Resta, D. J., Peng, B., Buchdunger, E., Ford, J. M., Lydon, N. B., Kantarjian, H., Capdeville, R., Ohno-Jones, S., and Sawyers, C. L. (2001) Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N. Engl. J. Med.* **344,** 1031–1037
2. O'Brien, S. G., Guilhot, F., Larson, R. A., Gathmann, I., Baccarani, M., Cervantes, F., Cornelissen, J. J., Fischer, T., Hochhaus, A., Hughes, T., Lechner, K., Nielsen, J. L., Rousselot, P., Reiffers, J., Saglio, G., Shepherd, J., Simonsson, B., Gratwohl, A., Goldman, J. M., Kantarjian, H., Taylor, K., Verhoef, G., Bolton, A. E., Capdeville, R., Druker, B. J., and the IRIS Investigators (2003) Imatinib compared with interferon and low-dose cytarabine for newly diagnosed chronic-phase chronic myeloid leukemia. *N. Engl. J. Med.* **348**, 994–1004
3. Demetri, G. D., von Mehren, M., Blanke, C. D., Van den Abbeele, A. D., Eisenberg, B., Roberts, P. J., Heinrich, M. C., Tuveson, D. A., Singer, S., Janicek, M., Fletcher, J. A., Silverman, S. G., Silberman, S. L., Capdeville, R., Kiese, B., Peng, B., Dimitrijevic, S., Druker, B. J., Corless, C., Fletcher, C. D., and

Joensuu, H. (2002) Efficacy and safety of imatinib mesylate in advanced gastrointestinal stromal tumors. *N. Engl. J. Med.* **347,** 472–480
4. Slamon, D. J., Leyland-Jones, B., Shak, S., Fuchs, H., Paton, V., Bajamonde, A., Fleming, T., Eiermann, W., Wolter, J., Pegram, M., Baselga, J., and Norton, L. (2001) Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N. Engl. J. Med.* **344,** 783–792
5. Romond, E. H., Perez, E. A., Bryant, J., Suman, V. J., Geyer, C. E., Jr., Davidson, N. E., Tan-Chiu, E., Martino, S., Paik, S., Kaufman, P. A., Swain, S. M., Pisansky, T. M., Fehrenbacher, L., Kutteh, L. A., Vogel, V. G., Visscher, D. W., Yothers, G., Jenkins, R. B., Brown, A. M., Dakhil, S. R., Mamounas, E. P., Lingle, W. L., Klein, P. M., Ingle, J. N., and Wolmark, N. (2005) Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer *N. Engl. J. Med.* **353,** 1673–1684
6. Piccart-Gebhart, M. J., Procter, M., Leyland-Jones, B., Goldhirsch, A., Untch, M., Smith, I., Gianni, L., Baselga, J., Bell, R., Jackisch, C., Cameron, D., Dowsett, M., Barrios, C. H., Steger, G., Huang, C.-S., Andersson, M., Inbar, M., Lichinitser, M., Lang, I., Nitz, U., Iwata, H., Thomssen, C., Lohrisch, C., Suter, T. M., Ruschoff, J., Suto, T., Greatorex, V., Ward, C., Straehle, C., McFadden, E., Dolci, M. S., Gelber, R. D., and the Herceptin Adjuvant (HERA) Trial Study Team (2005) Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *N. Engl. J. Med.* **353,** 1659–1672
7. Motzer, R. J., Michaelson, M. D., Redman, B. G., Hudes, G. R., Wilding, G., Figlin, R. A., Ginsberg, M. S., Kim, S. T., Baum, C. M., DePrimo, S. E., Li, J. Z., Bello, C. L., Theuer, C. P., George, D. J., and Rini, B. I. (2006) Activity of SU11248, a multitargeted inhibitor of vascular endothelial growth factor receptor and platelet-derived growth factor receptor, in patients with metastatic renal cell carcinoma. *J. Clin. Oncol.* **24,** 16–24
8. Ratain, M. J., Eisen, T., Stadler, W. M., Flaherty, K. T., Kaye, S. B., Rosner, G. L., Gore, M., Desai, A. A., Patnaik, A., Xiong, H. Q., Rowinsky, E., Abbruzzese, J. L., Xia, C., Simantov, R., Schwartz, B., and O'Dwyer, P. J. (2006) Phase II placebo-controlled randomized discontinuation trial of sorafenib in patients with metastatic renal cell carcinoma. *J. Clin. Oncol.* **24,** 2505–2512
9. Dumaz, N., Hayward, R., Martin, J., Ogilvie, L., Hedley, D., Curtin, J. A., Bastian, B. C., Springer, C., and Marais, R. (2006) In melanoma, RAS mutations are accompanied by switching signaling from BRaf to CRaf and disrupted cyclic AMP signaling. *Cancer Res.* **66,** 9483–9491
10. Davies, H., Bignell, G. R., Cox, C., Stephens, P., Edkins, S., Clegg, S., Teague, J., Woffendin, H., Garnett, M. J., Bottomley, W., Davis, N., Dicks, E., Ewing, R., Floyd, Y., Gray, K., Hall, S., Hawes, R., Hughes, J., Kosmidou, V., Menzies, A., Mould, C., Parker, A., Stevens, C., Watt, S., Hooper, S., Wilson, R., Jayatilake, H., Gusterson, B. A., Cooper, C., Shipley, J., Hargrave, D., Pritchard-Jones, K., Maitland, N., Chenevix-Trench, G., Riggins, G. J., Bigner, D. D., Palmieri, G., Cossu, A., Flanagan, A., Nicholson, A., Ho, J. W., Leung, S. Y., Yuen, S. T., Weber, B. L., Seigler, H. F., Darrow, T. L., Paterson, H., Marais, R., Marshall, C. J., Wooster, R., Stratton, M. R., and Futreal, P. A. (2002) Mutations of the BRaf gene in human cancer. *Nature* **417,** 949–954
11. Goydos, J. S., Mann, B., Kim, H. J., Gabriel, E. M., Alsina, J., Germino, F. J., Shih, W., and Gorski, D. H. (2005) Detection of B-RAF and N-RAS mutations in human melanoma. *J. Am. Coll. Surg.* **200,** 362–370
12. Goel, V. K., Lazar, A. J., Warneke, C. L., Redston, M. S., and Haluska, F. G. (2006) Examination of mutations in BRAF, NRAS, and PTEN in primary cutaneous melanoma. *J. Invest. Dermatol.* **126,** 154–160
13. Tsao, H., Goel, V., Wu, H., Yang, G., and Haluska, F. G. (2004) Genetic interaction between NRAS and BRAF mutations and PTEN/MMAC1 inactivation in melanoma. *J. Invest. Dermatol.* **122,** 337–341
14. Tsao, H., Zhang, X., Fowlkes, K., and Haluska, F. G. (2000) Relative reciprocity of NRAS and PTEN/MMAC1 alterations in cutaneous melanoma cell lines. *Cancer Res.* **60,** 1800–1804
15. Strumberg, D. (2005) Preclinical and clinical development of the oral multikinase inhibitor sorafenib in cancer treatment. *Drugs Today (Barc.)* **41,** 773–784
16. Ahmad, T., Marais, R., Pyle, L., James, B., Schwartz, B., Gore, M., and Eisen, T. (2004) BAY 43–9006 in patients with advanced melanoma: the Royal Marsden experience. *J. Clin. Oncol.* **22**(14S), 7506
17. Lorusso, P. M., Adjei, A. A., Varterasian, M., Gadgeel, S., Reid, J., Mitchell, D. Y., Hanson, L., DeLuca, P., Bruzek, L., Piens, J., Asbury, P., Van Becelaere, K., Herrera, R., Sebolt-Leopold, J., and Meyer, M. B. (2005) Phase I and pharmacodynamic study of the oral MEK inhibitor CI-1040 in patients with advanced malignancies. *J. Clin. Oncol.* **23,** 5281–5293
18. Ma'ayan, A., Jenkins, S. L., Neves, S., Hasseldine, A., Grace, E., Dubin-Thaler, B., Eungdamrong, N. J., Weng, G., Ram, P. T., Rice, J. J., Kershenbaum, A., Stolovitzky, G. A., Blitzer, R. D., and Iyengar, R. (2005) Formation of regulatory patterns during signal propagation in a Mammalian cellular network. *Science* **309,** 1078–1083
19. Bhalla, U. S., and Iyengar, R. (1999) Emergent properties of networks of biological signaling pathways. *Science* **283,** 381–387
20. Bhalla, U. S., Ram, P. T., and Iyengar, R. (2002) MAP kinase phosphatase as a locus of flexibility in a mitogen-activated protein kinase signaling network. *Science* **297,** 1018–1023
21. Kashtan, N., and Alon, U. (2005) Spontaneous evolution of modularity and network motifs. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 13773–13778
22. Garrington, T. P., and Johnson, G. L. (1999) Organization and regulation of mitogen-activated protein kinase signaling pathways. *Curr. Opin. Cell Biol.* **11,** 211–218
23. Vossler, M. R., Yao, H., York, R. D., Pan, M. G., Rim, C. S., and Stork, P. J. (1997) cAMP activates MAP kinase and Elk-1 through a B-Raf- and Rap1-dependent pathway. *Cell* **89,** 73–82
24. Busca, R., Abbe, P., Mantoux, F., Aberdam, E., Peyssonnaux, C., Eychene, A., Ortonne, J. P., and Ballotti, R. (2000) Ras mediates the cAMP-dependent activation of extracellular signal-regulated kinases (ERKs) in melanocytes. *EMBO J.* **19,** 2900–2910
25. Cook, S. J., and McCormick, F. (1993) Inhibition by cAMP of Ras-depenndent activation of Raf. *Science* **262,** 1069–1072
26. Wu, J., Dent, P., Jelinek, T., Wolfman, A., Weberr, M. J., and Sturgill, T. W. (1883) Inhibition of the EGF-activated MAP kinase signaling pathway by adenosine 3′,5′-monophosphae. *Science* **262,** 1065–1069
27. Dumaz, N., and Marais, R. (2005) Integrating signals between cAMP and the RAS/RAF/MEK/ERK signalling pathways. Based on the anniversary prize of the Gesellschaft fur Biochemie und Molekularbiologie Lecture delivered on 5 July 2003 at the Special FEBS Meeting in Brussels. *FEBS J.* **272,** 3491–3504
28. Dumaz, N., and Marais, R. (2003) Protein kinase A blocks Raf-1 activity by stimulating 14–3-3 binding and blocking Raf-1 interaction with Ras. *J. Biol. Chem.* **278,** 29819–29823
29. Brummer, T., Naegele, H., Reth, M., and Misawa, Y. (2003) Identification of novel ERK-mediated feedback phosphorylation sites at the C-terminus of B-Raf. *Oncogene* **22,** 8823–8834
30. Dougherty, M. K., Muller, J., Ritt, D. A., Zhou, M., Zhou, X. Z., Copeland, T. D., Conrads, T. P., Veenstra, T. D., Lu, K. P., and Morrison, D. K. (2005) Regulation of Raf-1 by direct feedback phosphorylation. *Mol. Cell* **17,** 215–224
31. Balan, V., Leicht, D. T., Zhu, J., Balan, K., Kaplun, A., Singh-Gupta, V., Qin, J., Ruan, H., Comb, M. J., and Tzivion, G. (2006) Identification of novel in vivo Raf-1 phosphorylation sites mediating positive feedback Raf-1 regulation by extracellular signal-regulated kinase. *Mol. Biol. Cell* **17,** 1141–1153
32. Jacobowitz, O., and Iyengar, R. (1994) Phorbol ester-induced stimulation and phosphorylation of adenylyl cyclase 2. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 10630–10634
33. Murphy, L. O., Smith, S., Chen, R. H., Fingar, D. C., and Blenis, J. (2002) Molecular interpretation of ERK signal duration by immediate early gene products. *Nat. Cell Biol.* **4,** 556–564
34. Cook, S. J., and McCormick, F. (1996) Kinetic and biochemical correlation between sustained p44ERK1 (44 kDa extracellular signal-regulated kinase 1) activation and lysophosphatidic acid-stimulated DNA synthesis in Rat-1 cells. *Biochem. J.* **320**(Pt 1), 237–245
35. Hoffman, A., Levchenko, A., Scott, M. L., and Baltimore, D. (2002) The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation. *Science* **298,** 1241–1245
36. Marshall, C. J. (1995) Specificity of receptor tyrosine kinase signaling: transient versus sustained extracellular signal-regulated kinase activation. *Cell* **80,** 179–185

37. Kholodenko, B. N. (2007) Untangling the signalling wires. *Nat. Cell Biol.* **9,** 247–249
38. Chin, L. (2003) The genetics of malignant melanoma: lessons from mouse and man. *Nat. Rev. Cancer* **3,** 559–570
39. Miller, A. J., and Mihm, M. C., Jr. (2006) Melanoma. *N. Engl. J. Med.* **355,** 51–65
40. Nesbit, M., Nesbit, H. K., Bennett, J., Andl, T., Hsu, M. Y., Dejesus, E., McBrian, M., Gupta, A. R., Eck, S. L., and Herlyn, M. (1999) Basic fibroblast growth factor induces a transformed phenotype in normal human melanocytes. *Oncogene* **18,** 6469–6476
41. Slominski, A., Tobin, D. J., Shibahara, S., and Wortsman, J. (2004) Melanin pigmentation in mammalian skin and its hormonal regulation. *Physiol. Rev.* **84,** 1155–1228
42. Landi, M. T., Bauer, J., Pfeiffer, R. M., Elder, D. E., Hulley, B., Minghetti, P., Calista, D., Kanetsky, P. A., Pinkel, D., and Bastian, B. C. (2006) MC1R germline variants confer risk for BRAF-mutant melanoma. *Science* **313,** 521–522
43. Bhalla, U. S. (2002) Mechanisms for temporal tuning and filtering by postsynaptic signaling pathways. *Biophys. J.* **83,** 740–752
44. Gao, L., Feng, Y., Bowers, R., Becker-Hapak, M., Gardner, J., Council, L., Linette, G., Zhao, H., and Cornelius, L. A. (2006) Ras-associated protein-1 regulates extracellular signal-regulated kinase activation and migration in melanoma cells: two processes important to melanoma tumorigenesis and metastasis. *Cancer Res.* **66,** 7880–7888
45. Alon, U. (2003) Biological networks: the tinkerer as an engineer. *Science* **301,** 1866–1867
46. Hornberg, J. J., Tijssen, M. R., and Lankelma, J. (2004) Synergistic activation of signalling to extracellular signal-regulated kinases 1 and 2 by epidermal growth factor and 4 beta-phorbol 12-myristate 13-acetate. *Eur. J. Biochem.* **271,** 3905–3913
47. Garnett, M. J., Rana, S., Paterson, H., Barford, D., and Marais, R. (2005) Wild-type and mutant B-RAF activate C-RAF through distinct mechanisms involving heterodimerization. *Mol. Cell* **20,** 963–969
48. Wan, P. T., Garnett, M. J., Roe, S. M., Lee, S., Niculescu-Duvaz, D., Good, V. M., Jones, C. M., Marshall, C. J., Springer, C. J., Barford, D., and Marais, R. (2004) Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF. *Cell* **116,** 855–867
49. Yamaguchi, T., Wallace, D. P., Magenheimer, B. S., Hempson, S. J., Grantham, J. J., and Calvet, J. P. (2004) Calcium restriction allows cAMP activation of the B-Raf/ERK pathway, switching cells to a cAMP-dependent growth-stimulated phenotype. *J. Biol. Chem.* **279,** 40419–40430
50. Sidovar, M. F., Kozlowski, P., Lee, J. W., Collins, M. A., He, Y., and Graves, L. M. (2000) Phosphorylation of serine 43 is not required for inhibition of c-Raf kinase by the cAMP-dependent protein kinase. *J. Biol. Chem.* **275,** 28688–28694
51. Santos, S. D., Verveer, P. J., and Bastiaens, P. I. (2007) Growth factor-induced MAPK network topology shapes Erk response determining PC-12 cell fate. *Nat. Cell Biol.* **9,** 324–330
52. Sasagawa, S., Ozaki, Y., Fujita, K., and Kuroda, S. (2005) Prediction and validation of the distinct dynamics of transient and sustained ERK activation. *Nat. Cell Biol.* **7,** 365–373
53. Robubi, A., Mueller, T., Fueller, J., Hekman, M., Rapp, U. R., and Dandekar, T. (2005) B-Raf and C-Raf signaling investigated in a simplified model of the mitogenic kinase cascade. *Biol. Chem.* **386,** 1165–1171
54. Satyamoorthy, K., Li, G., Gerrero, M. R., Brose, M. S., Volpe, P., Weber, B. L., Van Belle, P., Elder, D. E., and Herlyn, M. (2003) Constitutive mitogen-activated protein kinase activation in melanoma is mediated by both BRAF mutations and autocrine growth factor stimulation. *Cancer Res.* **63,** 756–759
55. Sun, S. Y., Rosenberg, L. M., Wang, X., Zhou, Z., Yue, P., Fu, H., and Khuri, F. R. (2005) Activation of Akt and eIF4E survival pathways by rapamycin-mediated mammalian target of rapamycin inhibition. *Cancer Res.* **65,** 7052–7058

Received for publication August 20, 2007.
Accepted for publication November 29, 2007.

Rapidly exploring structural and dynamic properties of signaling networks using PathwayOracle

Derek Ruths*[1], Luay Nakhleh[1]and Prahlad T. Ram[2]

[1]Department of Computer Science, Rice University, Houston, Texas, USA
[2]Department of System Biology, University of Texas MD Anderson Cancer Center, Houston, Texas, USA

Email:
DR: druths@cs.rice.edu
LN: nakhleh@cs.rice.edu
PTR: pram@mdanderson.org ;

*Corresponding author

Abstract

Background: In systems biology the experimentalist is presented with a selection of software for analyzing dynamic properties of signaling networks. These tools either assume that the network is in steady-state or require highly parameterized models of the network of interest. For biologists interested in assessing how signal propagates through a network under specific conditions, the first class of methods does not provide sufficiently detailed results and the second class requires models which may not be easily and accurately constructed. A tool that is able to characterize the dynamics of a signaling network using an unparameterized model of the network would allow biologists to quickly obtain insights into a signaling network's behavior.

Results: We introduce $PathwayOracle$, an integrated suite of software tools for computationally inferring and analyzing structural and dynamic properties of a signaling network. The feature which differentiates $PathwayOracle$ from other tools is a method that can predict the response of a signaling network to various experimental conditions and stimuli using only the connectivity of the signaling network. Thus signaling models are relatively easy to build. The method allows for tracking signal flow in a network and comparison of signal flows under different experimental conditions. In addition, $PathwayOracle$ includes tools for the enumeration and visualization of coherent and incoherent signaling paths between proteins, and for experimental analysis—loading and superimposing experimental data, such as microarray intensities, on the network model.

Conclusions: *PathwayOracle* provides an integrated environment in which both structural and dynamic analysis of a signaling network can be quickly conducted and visualized along side experimental results. By using the signaling network connectivity, analyses and predictions can be performed quickly using relatively easily constructed signaling network models. The application has been developed in Python and is designed to be easily extensible by groups interested in adding new or extending existing features. *PathwayOracle* is freely available for download and use.

Background

Reconstructing cellular signaling networks and understanding how they work are major endeavors in cell biology. The scale and complexity of these networks, however, render their analysis using experimental biology approaches alone very challenging. As a result, computational methods have been developed and combined with experimental biology approaches, producing powerful tools for the analysis of these networks. These tools aid biologists in interpreting existing experimental findings, evaluating hypotheses, enumerating possible biological behaviors, and, ultimately, in quickly designing experiments that maximize the amount of useful information gained. By assisting biologists in maximizing the amount of information obtained from their experiments through improved experimental design and more thorough analysis of results, computational tools increase the pace of scientific discovery.

Biological network analysis can generally be classified as either *structural* or *dynamic* [1]. Structural analysis provides insights into global properties of the network, among them decomposition of the network into functional modules (e.g., [2]), enumeration of signaling paths connecting arbitrary protein pairs (e.g., [3–5]), and the identification of key pathways that determine the behavior of the network (e.g., [2,6–10]). Dynamic methods, on the other hand, simulate the actual propagation of signals through a network by predicting the changes in the concentration of signaling proteins over time. These predictions will be of varying degrees of resolution and accuracy, depending largely on the accuracy and level of detail of the model from which they are produced.

The prevailing methods for dynamic analysis involve systems of ordinary differential equations (ODEs) [11,12]. These approaches require kinetic parameters for the individual biochemical reactions

involved in the signaling process. This requirement often poses a significant hurdle for researchers as the numerical values of such parameters are difficult to obtain and may be the object of the researcher's project in the first place. In [13], we presented a novel signaling network simulation method which uses a non-parametric Petri net model of network to predict the signal flow under various experimental conditions. Our simulation method uses a novel technique to approximate the interaction speeds and predicts the qualitative behavior of the signaling network dynamics.

The advantage of our method over ODEs is the wide availability of connectivity-based models of signaling networks, and the relative speed with which they can be constructed. Numerous databases exist which catalog known signaling interactions (e.g., [14–16]). Thus, the existence and type (activating or inhibition) of an interaction can often be inferred directly from literature and/or these databases. This presents a stark contrast to the kinetic parameters required by ODEs, the numerical values for many of which must be determined experimentally for each experimental condition and cell line of interest [2].

In this paper, we present the software tool *PathwayOracle*, an integrated environment for connectivity-based structural and dynamic analysis of signaling networks, supporting

- visualization of signaling network connectivity;
- two versions of the simulation method described in [13] where
  - the first allows prediction of signal flow through a given network for a specific experimental condition, and
  - the second predicts the difference in signal flow through a given network induced by two different experimental conditions;
- enumeration of the paths connecting arbitrary pairs of nodes in the network; and
- visualization of experimental concentration data on the signaling network display.

In future releases we plan on expanding capabilities in all three areas of analysis—dynamic, structural, and experimental—with a focus on providing effective ways of integrating results from each together. *PathwayOracle* has been designed in a modular fashion in order to facilitate extension of existing capabilities and the addition of new features.

Since PathwayOracle's most distinctive analytical capability involves the signaling Petri net simulator, a new dynamic analysis technique for signaling networks, we first provide an overview of the signaling Petri

net modeling approach. Then in subsequent sections, we focus on PathwayOracle and explain the architecture and core concepts underlying the tool and then examine the individual features, how they can be used, and how they compare to existing tools.

The Signaling Petri Net Simulator

Petri nets provide a graphical and executable model of processes in which information or material flows among a series of places or entities [17]. A Petri net consists of places, transitions, and tokens (see Figure 1). Quantities of tokens are assigned to individual places. This assignment is called a *marking*. As Figure 1 illustrates, the network flow is modeled by the reassignment of tokens to individual places in the Petri net in response to transition firings.

A signaling Petri net is an extension of the Petri net formalism to model a signaling network. Places are signaling proteins and transitions implement directed protein interactions; each transition models the effect of a source protein on a target protein. The marking of (number of tokens in) protein p at time t is interpreted as the activity-level of that protein—the number of activated molecules of that type. Figure 2 shows the correspondence between a signaling network and a signaling Petri net model.

The signaling Petri net simulator models signal flow as the pattern of token accumulation and dissipation within proteins over time in the Petri net. Through transition firings, the source can influence the marking of (the number of tokens assigned to) the target, modeling the way that signals propagate through protein interactions in cellular signaling networks.

In order to overcome the issue of modeling reaction rates in the network, signaling dynamics are simulated by executing the signaling Petri net (SPN) for a set number of steps (called a *run*) multiple times, each time beginning at the same initial marking. For each run, the individual signaling rates are simulated via generation of random orders of transition firings (interaction occurrences). When the results of a large enough number of runs are averaged together, we find that the change in distribution of tokens in the network correlate with experimentally measured changes in the activity-levels of individual proteins in the underlying signaling network. In essence, the tokenized activity-levels computed by our method should be taken as abstract quantities whose changes over time correlate to changes that occur in the amounts of active proteins present in the cell. It is worth noting that some of the most widely used experimental techniques for protein quantification—western blots and microarrays—also yield results that are treated as indications, but not exact measurements, of protein activity-levels within the cell. Thus in some respects, the predictions returned by our SPN-based simulator can be interpreted like the results of a western blot or

microarray experiment looking at changes relative to "control".

During a simulation run, the simulator imposes a strict ordering on transition firing such that it creates a two-time scale simulation. The smaller time scale is discretized as the firing of a single transition. This unit is referred to as the *firing* time scale. Firing steps are nested within a larger time scale, called time *blocks*, within which each transition is fired exactly once. The values returned by the simulator are the averaged token-counts for each protein at each time-block (across all runs).

Figure 3 provides a small example of a simulation run whose duration is two time blocks. As mentioned previously, within a given time block, each transition fires exactly once. Thus, in the table (Figure 3(c)), there is one column for each transition in each time block. The ordering of the transitions is shuffled in each time block in order to sample a different set of signaling rates within the networks.

In the first time block, transition t_2 fires first: it reads 2 tokens out of *Grb2* and places 2 additional tokens in *Ras*. Transition t_1 fires second, reading 3 tokens out of *Grb2*. Transition t_3 is evaluated last. The final marking for the network, highlighted as the red column in block 1 is used by the simulator as the marking for that block when averaging across runs.

At the conclusion of block 2, compare the values highlighted in red in the Initial column and at the end of both blocks. Note how the distribution of tokens have changed over the course of the simulation. *Grb2* has the same number of tokens, implying that its activity-level has remained unchanged—this is consistent with the signaling network since no activating or inhibiting edges affect it in the model. *AKT*'s token-count has risen, consistent with the fact that it is only activated in the signaling network. *Ras*'s token-count has fallen which is one plausible behavior of the system since it is activated by *Grb2*, but inhibited by *AKT*.

Implementation

PathwayOracle is written in Python [18]. The user experience is oriented around visualization of and interaction with three main types of data: the signaling network, markings, and paths. At any given time, one signaling network is open, which is the basis for all analyses. Any simulation or concentration data is loaded and inspected as markings. Currently all static analyses revolve around paths, which are the third data type. In the following subsections, these individual data types and the user interfaces to them are discussed in more detail.

The Signaling Network Model

While the implementation of our methods use the signaling Petri net model discussed in an earlier section of this paper, we provide a simpler and more convenient representation of the network to the user which omits the internal topology of the transitions and allows the user to specify interactions simply as either activating or inhibiting. Thus, for the remainder of this paper we use the following definition of the signaling network which is consistent with the experience the user will have when working with *PathwayOracle*. The signaling network connectivity is a directed graph $G = (V, E)$ where

- V is the set of nodes, which are signaling proteins and complexes (hereafter referred to collectively as *signaling nodes*) and
- E is the set of edges, which are signaling interactions. Each edge is of one of two types: $u \rightarrow v$ for activation and $u \dashv v$ for inhibition.

Within *PathwayOracle*, each signaling node has a name, unique within the network. A signaling edge has no properties besides its type and is only defined by its *source* and *target*.

In order to facilitate the rapid construction of such signaling network models, we devised a file format called the *Connectivity Format*. It is capable of expressing both general networks as well as paths. When representing a network in the format, as shown in the example in Figure 4(b), one signaling interaction is written on a line with the format

$$u \text{ -> } v \qquad \text{or} \qquad u \text{ -| } v$$

where u is the name of the source signaling node and v is the name of the target signaling node. Each node is taken to represent the active form of the protein it is named for. Thus, from the example above, the interaction PI-3-K→AKT means that the active form of PI-3-K increases the activity-level of AKT whereas the interaction PTEN⊣AKT means that the active form of PTEN decreases the activity-level of AKT. While these types of unparameterized relationships can be represented in SBML, SBML was designed for encoding much more information than just connectivity [19]. As a result, we deemed it appropriate to design a more concise format for our purposes. However, in a future release, *PathwayOracle* will support loading and saving in the SBML format.

At a given point in time, only one signaling network can be open in *PathwayOracle*. The main window displays a graphical representation of the network. The layout of the network can be modified by dragging

nodes or by *shift*-clicking on edges to create, remove, or move waypoints. These layouts can be saved with the network and loaded again.

Signaling Network Markings

In signaling networks, signal flow is measured and quantified as the fluctuation of concentrations of various forms of signaling proteins over time. In *PathwayOracle*, we model concentrations using the concept of a network *marking*, which was adapted from Petri nets in which it was first used [9].

Markings

In *PathwayOracle*, a marking, μ is an assignment of real values to the nodes of a signaling network such that every signaling node receives a value. Earlier, the concept of a marking was introduced as the assignment of tokens to protein places in the signaling Petri net. In a signaling Petri net, tokens are discrete. In *PathwayOracle*, a marking is an average of the markings from many independent simulation runs, which gives rise to the real, rather than integral values, assigned by the marking.
As discussed earlier, the value of the marking of a signaling node, $\mu(v)$, can be interpreted as an estimate of the concentration or change in concentration of the active form of the signaling protein v (we call the amount of the active form of the signaling protein its *activity-level*). The two different versions of the simulator generate markings with these different meanings. The first simulator predicts the signal flow due to an experimental condition and generates markings whose values are taken to represent the actual activity-level of signaling protein present over the assumed basal levels. The second version of the simulator predicts the difference in signaling due to changing experimental conditions. The values assigned by markings produced by this simulator correspond to the *change* in the activity-level of the protein induced by the change in experimental condition. This will be discussed further in the Results and Discussion section.

Marking Series

In order to model signal *flow*, a single marking is not enough since it only provides a single snapshot of concentrations throughout the network. A *marking series* is an sequence of markings, $(\mu_1, \mu_2, ..., \mu_T)$ in which the marking μ_t is a snapshot of the concentration distribution at time step t. Thus, it is possible to see how the activity-level of protein v changed by plotting the values $\mu_1(v), \mu_2(v), ..., \mu_T(v)$. *PathwayOracle* provides the ability to do this.

PathwayOracle supports loading a marking series dataset from *comma-separated value* (*.csv*) files. As shown in Figure 5(a), the file has a header row which specifies, for each column, the name of the molecule whose concentration values will appear in that column. Each subsequent row contains the value assignments for a marking: the second row contains the marking for time step 1, the third row contains the marking for time step 2, and so on.

Marking Groups

In many experiments, the activity-level of various proteins are sampled at different time points and under different experimental conditions. Since the *marking series* is not able to represent changes due to different experimental conditions, we introduced the more general concept of a *marking group* in which each marking can correspond to an arbitrary activity-level distribution. Each marking is given a descriptive label that can be used to identify the conditions under which the activity-level was sampled.
Like the marking series, a marking group is loaded from a *.csv* file. However, unlike the marking series in which each row corresponds to a time step, in the marking group, each row corresponds to an independent marking (experimental condition). As shown in Figure 5(b), the first row is a header row specifying the molecule names for each column, the first column specifies the names for the individual markings (experimental conditions).

The Marking Manager

PathwayOracle includes a specific user-interface, the *Marking Manager*, designed to manage the three different types of markings. The Marking Manager provides a central interface within which it is possible to view all markings loaded and inspect them in ways that are relevant to their type (marking, marking series, or marking group). The specific ways in which markings can be inspected will be discussed further in the *Results* section.

Signaling Paths

The current structural analysis capabilities available in *PathwayOracle* allow inspection of signaling paths within the network. A signaling path p is a sequence of nodes, $(v_1, v_2, ..., v_k)$ where $v_i \in V \ \forall 1 \leq i \leq k$, and $(v_i, v_i + 1) \in E \ \forall 1 \leq i < k$. In this case, we say that node v_1 is the *source* of path p, and node v_k is the *target* of p. Given a path, a variety of statistics may be of interest to the user. Additionally, it may be useful to view the path within the larger network. *PathwayOracle* provides these capabilites which will be

discussed in the Results and Discussion section.

Sets of paths can be saved to a file and loaded back into a session. Like networks, paths are also stored in the Connectivity Format. When representing a set of paths, as shown in Figure 6, the full node names and the edge types are written so that all path information is directly available within the file itself. One line contains one path.

Results

PathwayOracle provides a variety of tools for analyzing the structural and dynamic properties of a signaling network based on its connectivity. While its main differentiating feature is the ability to predict signal flow through a network using only the connectivity of the signaling network, *PathwayOracle* also provides the ability to visualize the network, analyze its connectivity, and inspect concentration-based experimental data.

With the exception of the signaling Petri net simulator, PathwayOracle's features can be found in various combinations in other tools. Figure 7 provides a matrix of the features and capabilities of several tools most commonly-used for signaling network analysis. While other tools support a variety of simulation techniques, PathwayOracle, alone, provides non-parameterized simulation capabilities. It is worth noting that the commercial software package CellIllustrator [20] provides Petri net-based simulation capabilities. The difference between CellIllustrator and PathwayOracle Petri net approaches is the extensive set of kinetic parameters required by CellIllustrator in order to simulate a biological system. In this regard, hybrid functional Petri nets, the underlying technology used by CellIllustrator, are not significantly different from ODEs.

Another important distinguishing characteristic of PathwayOracle is the combination of features that it supports. Biological network analysis is a multi-faceted process that may involve structural, dynamic, and data analysis in parallel. Whereas other tools tend to focus on one or two of these general areas of analysis, we considered it important for PathwayOracle to incorporate all three in order to provide the researcher a single environment in which all their analysis could be done. In future releases we plan to increase PathwayOracle's support for all three of these directions of investigation: structural, dynamic, and data analysis.

In the remainder of this section, we discuss the features currently available in PathwayOracle.

Network Visualization

As in many other computational analysis tools for signaling networks (e.g., [20, 21]), an interactive graphical representation of the signaling network connectivity is at the center of the *PathwayOracle* interface. The main window provides a visualization of the signaling network connectivity. This visualization interface allows the user to edit the layout of the network by clicking on and dragging nodes and by *shift*-clicking on edges to create, remove, or move waypoints. Waypoints are points that lie on an edge. Holding down *shift* will display all edge waypoints. Existing waypoints can be dragged to change the path that an edge follows. Right-clicking on a waypoint will remove it. Left-clicking on a straight segment of the edge will create a new waypoint.

The network visualization also provides a view onto which path and experimental data analysis may be mapped. As will be discussed in subsequent sections, selected paths may be highlighted in this view and markings from experiments can set the colorings of individual nodes.

Network Signal Flow Simulation

The main feature differentiating *PathwayOracle* from other tools, such as CellDesigner [20] and COPASI [22], is its ability to simulate signal flow using an unparameterized signaling network model. Simulations can be performed in two different ways. In the first (*Single Simulation*), the simulator predicts the signal flow through the network for a specific experimental condition. In the second (*Differential Simulation*), the simulator predicts the difference in signal flow due to two different experimental conditions on the same network. These simulation methods themselves are described in [13]. Here we focus on how simulations are configured, run, and analyzed.

Whereas the consensus networks typically represent the connectivity in normal cells, many experiments are conducted on abnormal cells in which oncogenic mutations, gene knockous, and pharmacological inhibitors have altered the behavior of various signaling nodes in the network. In *PathwayOracle* users can model these cell- and experiment-specific conditions by specifying each signaling node as either *High*, *Low*, or *Free*. The *High* state models any condition under which a protein's activity-level is held high for the duration of the experiment. This may be due to external stimulation or a known mutation in the protein that makes it constitutively active, for example. Similarly, a *Low* state models any phenomenon that forces a protein to have a persistently suppressed activity-level. This may be due to mutations that render the protein inactive, gene knockouts, or pharmacological inhibitors that force the activity-level of the protein low. In general, most signaling nodes will be *Free*, which means that their activity-level is unconstrained

throughout the simulation. Only those nodes designated as *High* or *Low* will have their activity-level fixed for the duration of the simulation.

In order for a protein to be held high during the simulation, it is necessary to indicate the initial activity-level that the protein will be elevated to. This is done by specifying the number of tokens that the protein will receive. Since a protein with a *High* state cannot be inhibited (even if inhibitory edges target it in the actual network), the protein's activity level will never fall below this initial value. The initial value for a High protein is indicated by placing it in parentheses next to the protein's name, as shown in Figure 8.

Two other parameters that must be specified for a simulation are:

- the number of simulation runs to perform and
- the number of time blocks

The number of runs sets the number of independent simulations whose time block markings are averaged together to yield the overall simulation markings. In general, using more runs is a tradeoff between reliability of the results and simulation speed. In practice, the number of runs needed is dependent on the signaling network model and should be selected by observing the reproducability of the simulation results. An appropriate number of iterations will be large enough so that for a given experimental condition, the results are very similar across multiple simulations.

The time block, as discussed earlier, is a fundamental unit of time in the simulator. The appropriate number of time blocks for which to simulate will vary depending on the size of the signaling network and the scale of the network behavior of interest. Generally it should be selected by running simulations for a variety of time block values and determining which yields the most biologically reasonable activity-level changes for a known protein. While this is a manual process in the current version of *PathwayOracle*, we are investigating automated methods for estimating the number of time blocks by training against experimental time series data.

In *PathwayOracle*, the setup window for the *Single Simulation* (see Figure 8(a)) prompts the user for a single experimental condition. The setup window for the *Differential Simulation* (see Figure 8(b)) prompts the user for two experimental conditions. Both simulators produce a marking series. The tokenized simulation marking series corresponds to the activity-level time series predicted for the specified experimental condition. The differential simulation marking series corresponds to the change in activity-levels over time produced by switching from experimental condition 2 to experimental condition 1. The marking series produced by a simulation can be accessed through the Marking Manager. Choosing to

inspect a marking series will present the user with a blank plot. By selecting signaling nodes, the plot is populated by the marking series values for individual nodes over time, as shown in Figure 8(c).
While this plot generation capability exists in many other dynamic simulation tools, the simplicity of the model used for simulation and the speed with which a simulation runs set *PathwayOracle* apart from other tools which require specification of the numerical values of kinetic parameters for each reaction in the network of interest (e.g., [20, 22]). *PathwayOracle*, because of its novel approach, does not have such requirements. It is worth noting, however, where *PathwayOracle* provides approximations of signal flow, an ODE generates the actual concentration changes using extremely detailed and accurate models of the underlying biochemistry. The simulators in *PathwayOracle* provide an attractive, time- and resource-saving alternative this more exhaustively parameterized techniques. In particular, *PathwayOracle's* features will benefit researchers interested in quickly assessing characteristics of signal flow in their network.
For some networks, biologists will have partial knowledge of kinetic parameters or of other biological details which the signaling Petri net model does not, at present, consider. By integrating this knowledge into the simulator, it may be possible to improve the simulator's predictions. We identify this as a direction for future investigation. As the signaling Petri net simulator is extended, these new capabilities will be incorporated in future releases of *PathwayOracle*.

Signaling Path Analysis

The use of the simulators and plotting tools allows the user to observe trends in the activity-level of individual signaling nodes over time. Since the activity-level of a node is determined by the activity-level of other nodes in the network, the activity-level time series of a node may be explained by changes in the activity-level history of nodes upstream of it. In order to investigate such indirect interactions, it is useful to enumerate all the paths leading from a specific protein to the protein of interest. *PathwayOracle* provides this capability. Additionally, it provides various statistics on the set of paths linking two signaling nodes as well as a classification of the effect of each path as either *coherent* or *incoherent* (e.g. [23]).
A coherent path is a directed series of interactions that leads from x to y such that an increase in the activity-level of x causes an increase in the activity of y and a decrease in the activity-level of x causes a decrease in the activity-level of y. An incoherent path is a directed series of interactions leading from x to y such that an increase in the activity-level of x causes a *d*ecrease in the activity-level of y and a decrease in the activity-level of x causes a *i*ncrease in the activity-level of y. It is possible to classify a path p as either coherent or incoherent by counting the number of inhibitory edges along p. *A path with an even*

number of inhibitory edges is coherent; a path with an odd number of inhibitory edges is incoherent [5]. This logic is assumed in *PathwayOracle*. All simple paths (paths without loops) connecting two specified signaling nodes are enumerated by an exhaustive depth-first search. These paths then are classified as either coherent or incoherent, and presented to the user for further inspection in a window similar to the one shown in Figure 9(a). When a path is selected in the results window, it is highlighted in the main window, allowing the user to evaluate it within the context of the complete network (see Figure 9(b)).

Experimental Data Analysis

A model of the connectivity of a signaling network makes it possible to identify components of the model that are inconsistent with experimental data or visa versa. *PathwayOracle* enables this kind of analysis by allowing users to load experimental concentration data and visualize it both as a heatmap (see Figure 10(a)) or superimposed on the network view (see Figure 10(b)). Several other software tools provide similar capabilities (e.g., [21]). In *PathwayOracle*, experimental concentration data is loaded as a marking group in which a single marking corresponds to a condition for which concentrations were sampled. Figure 10(a) shows a marking group with 24 conditions (rows). The concentration of seven signaling proteins were sampled for each condition. This is the heatmap view for the marking group. When a specific marking in the group is selected, the colors for that marking are applied to the network view. This is particularly useful when assessing whether the experimental data is consistent with the interactions in the model. In Figure 10, the MDA231-B-DMSO2 marking has been superimposed on the network. We can see that RSK has a relatively low concentration despite the high concentration of MAPK. Given that, in the model, RSK is activated by MAPK, this combination of activity-levels seems unlikely to occur. Such an inconsistency suggests that there may be other signaling interactions contributing to the overall activity-level of RSK. Such an insight can help a researcher quickly identify areas where the model or experimental results need to be re-evaluated or improved.

Future Directions

Our goal is to develop *PathwayOracle* into an integrated and expansive suite of tools that allow the biologist to extract as much information as possible from models of signaling network connectivity and experimental data relating to those models. We consider future directions for *PathwayOracle* to fall into several categories: network construction, network augmentation, experimental and computational analysis integration, and architecture.

One of the benefits of working with connectivity models of signaling networks is the abundance of databases and other online resources that publish connectivity-level data. Future versions of *PathwayOracle* will have support for querying such databases for connectivity components and, ultimately, for automated connectivity construction based on a set of signaling nodes specified by the user.

Analysis of network connectivity and topology is increasingly relevant to biological research. We intend to expand PathwayOracle's structural analysis features to include the ability to search for and identify motifs in the signaling networks.

Network connectivity can also be inferred from experimental data, which provides another direction for research and development. By using experimental results to identify inconsistencies between experimental results and the current network model, it may be possible for *PathwayOracle* to augment the network with new connectivity based on hints supplied by experimental results. At present only experimental concentration data is supported. However, as experiments produce more information beyond concentration profiles of signaling nodes, we plan to expand the experimental data that *PathwayOracle* can load, visualize, and use as part of network analyses.

Experimental results can also provide computational analysis methods information that can improve their final predictions or decompositions. Taking advantage of the additional, potentially obfuscated, information present in experimental results to improve the results returned by computational tools is a major goal for future versions of *PathwayOracle*.

A longer term direction for *PathwayOracle* is the integration of transcriptional and metabolic network analysis. In the biological systems of interest, the behavior of any one of these networks is dependent on the characteristics of the other two. As a result, developing a complete understanding of signaling, transcriptional regulation, or metabolism depends in part on integrating knowledge from the others.

Finally, an ongoing priority in the design of *PathwayOracle* is its role as an open platform for the development and deployment of new analytical capabilities by other groups. Currently *PathwayOracle* employes a modular architecture that facilitates easy integration of new functionality. However, in future releases we plan to expose a plugin interface which will make it easier to developers and researchers to develop and deploy tools within PathwayOracle.

Conclusions

PathwayOracle is an integrated software environment in which biologists may conduct structural and dynamic analysis of signaling networks of interest. *PathwayOracle* is distinguished from other tools in the

field of systems biology by its ability to predict the signal flow through a network using a simplified, connectivity-based model of the signaling network. Simulations are fast and, based on a published study, predictors of signal propagation. This novel simulation capability, combined with support for structural analysis of connectivity between pairs of proteins and for analysis of certain kinds of experimental data make *PathwayOracle* a powerful asset in the experimentalist's endeavor to gain a more complete understanding of the cellular signaling network.

Availability and requirements

Project name: PathwayOracle

Project home page: http://bioinfo.cs.rice.edu/pathwayoracle

Operating system(s): Platform independent

Programming language: Python

Other requirements: Python 2.4 or higher

License: GNU GPL

Any restrictions to use by non-academics: None

Funding

D. R. and L. N. are supported in part by a Seed Grant awarded to L.N. from the Gulf Coast Center for Computational Cancer Research, funded by John and Ann Doerr Fund for Computational Biomedicine. P.T.R. is supported in part by a Department of Defense grant BC044268.

References

1. Papin JA, Hunter T, Palsson BO, Subramaniam S: **Reconstruction of cellular signalling networks and analysis of their properties**. *Nature Reviews Molecular Cell Biology* 2005, **6**:99–111.
2. Sackmann A, Heiner M, Koch I: **Application of Petri net based analysis techniques to signal transduction pathways**. *BMC Bioinformatics* 2006, **7**:482.
3. Ruths D, Tseng JT, Nakhleh L, Ram PT: **De novo Signaling Pathway Predictions based on Protein-Protein Interaction, Targeted Therapy, and Protein Microarray Analysis**. In *Proceedings of the RECOMB Satellite Workshop on Systems Biology and Proteomics, Lecture Notes in Bioinformatics (LNBI #4466)* 2007:62–72.
4. Ruths D, Nakhleh L, Iyengar MS, Reddy SAG, Ram PT: **Graph-theoretic Hypothesis Generation in Biological Signaling Networks**. *Journal of Computational Biology* 2006, **13**(9):1546–1557.

5. Klamt S, Saez-Rodriguez J, Lindquist JA, Simeoni L, Gilles ED: **A methodology for the structural and functional analysis of signaling and regulatory networks**. *BMC Bioinformatics* 2006, **6**:56.
6. Papin JA, Palsson BO: **The JAK-STAT signaling network in the human B-cell: an extreme signaling pathway analysis**. *Biophysical Journal* 2004, **87**:37–46.
7. Papin JA, Price ND, Wiback SJ, Fell DA, Palsson BO: **Metabolic pathways in the post-genomic era**. *TRENDS in Biochemical Sciences* 2003, **28**(5):250–258.
8. Schilling CH, Letscher D, Palsson BO: **Theory for the Systemic Definition of Metabolic Pathways and their use in Interpreting Metabolic Function from a Pathway-Oriented Perspective**. *Journal of Theoretical Biology* 2000, **203**:229–248.
9. Peleg M, Rubin D, Altman RB: **Using Petri Net Tools to Study Properties and Dynamics of Biological Systems**. *Journal of the American Medical Informatics Association* 2005, **12**(2):181–199.
10. Chaouiya C: **Petri net modelling of biological networks**. *Briefings in Bioinformatics* 2007, **8**(4):210–219.
11. Eungdamrong NJ, Iyengar R: **Modeling cell signaling networks**. *Biology of the Cell* 2004, **96**(5):355–362.
12. Eungdamrong NJ, Iyengar R: **Computational Approaches for modeling regulatory cellular networks**. *Trends Cell Biology* 2004, **14**(12):661–669.
13. Ruths D, Muller M, Tseng JT, Nakhleh L, Ram PT: **The Signaling Petri Net-based Simulator: A Non-Parametric Strategy for Characterizing the Dynamics of Cell-Specific Signaling Networks**. *PLoS Computational Biology* 2008, **4**(2):e1000005.
14. Kanehisa M, Goto S: **KEGG: Kyoto Encyclopedia of Genes and Genomes**. *Nucleic Acids Research* 2000, **28**(15):27–30.
15. **The Cancer Cell Map** [http://cancer.cellmap.org].
16. Thomas PD, Campbell MJ, Kejariwal A, Mi J, Karlak B, Daverman R, Diemer K, Muruganujan A, Narechania A: **PANTHER: a library of protein families and subfamilies indexed by function**. *Genome Research* 2003, **13**:2129–2141.
17. David R, Alla H: *Discrete, Continuous, and Hybrid Petri Nets*. Springer 2005.
18. **Official Website for Python Programming Language** [http://www.python.org].
19. Hucka M, Finney A, Sauro HM, H B, Doyle J, Kitano H, Arkin A, Bornstein B, Bray D, Cornish-Bowden A, Cuellar A, Dronov S, Gilles E, Ginkel M, Gor V, Goryanin I, Hedley W, Hodgman T, Hofmeyr J, Hunter P, Juty N, Kasberger J, Kremling A, Kummer U, Le Novere N, Loew L, Lucio D, Mendes P, Minch E, Mjolsness E, Nakayama Y, Nelson M, Nielsen P, Sakurada T, Schaff J, Shapiro B, Shimizu T, Spence H, Stelling J, Takahashi K, Tomita M, Wagner J, Wang J: **The Systems Biology Markup Language (SBML): A medium for representation and exchange of biochemical network models**. *Bioinformatics* 2003, **19**(4):524–531.
20. Funahashi A, Tanimura N, Morohashi M, Kitano H: **CellDesigner: a process diagram editor for gene-regulatory and biochemical networks**. *BIOSILICO* 2003, **1**:159–162.
21. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks**. *Genome Research* 2003, **13**(11):2498–504.
22. Hoops S, Sahle S, Gauges R, Lee C, Pahle J, Simus N, Singhal M, Xu L, Mendes P, Kummer U: **COPASI—a COmplex PAthway SImulator**. *Bioinformatics* 2006, **22**:3067–74.
23. Alon U: *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Mathematical and Computational Biology Series, Chapman & Hall/CRC 2007.
24. Nagasaki M, Doi A, Matsuno H, Miyano S: **Genomic Object Net: I. A platform for modelling and simulating biopathways**. *Applied Bioinformatics* 2003, **2**(3):181–184.
25. Klamt S, Saez-Rodriguez J, Gilles ED: **Structural and functional analysis of cellular networks with CellNetAnalyzer**. *BMC Systems Biology* 2007, **1**:2.
26. Schmidt H, Jirstand M: **Systems Biology Toolbox for MATLAB: A computational platform for research in Systems Biology**. *Bioinformatics* 2006, **22**(4):514–515.

Figure legends

Figure 1: **An example of how tokens move among places.** In a Petri net, quantities of tokens are assigned to places. In (a), three tokens are assigned to place p_A and zero tokens are assigned to place p_B. The two places are connected by a transition, t_1. The arcs in and out of t_1 indicate the direction in which tokens move. When t_1 fires, it moves some number of tokens from p_A and puts them in p_B. In (b), transition t_1 has fired and moved two tokens from p_A to p_B.

Figure 2: **An example signaling network and its corresponding Petri net.** An example signaling network (a) and its corresponding Petri net (b). Each signaling protein in the network, A, B, and C, is designated as a place p_A, p_B, and p_C. A signaling interaction becomes a transition node and its input and output arcs. Note that the connectivity for an activating edge differs from that of an inhibitory edge.

Figure 3: **An example signaling Petri net simulation.** (a) is the signaling network being simulated. (b) is the signaling Petri net that models that signaling Petri net. The table in (c) provides the markings for the Petri net over the course of a simulation run whose duration is two time blocks. The proteins are given the initial marking shown in the *Initial* column. Each subsequent column corresponds to a single time step during which one transition fired, producing a new marking of the network. The bold number in each column indicates which protein's marking was affected by the transition that fired in that time step. The red columns—always the last time step in the block—highlight the markings whose values would be averaged and used as part of the final result. These red columns are the sources of the markings that *PathwayOracle* reports.

Figure 4: **An example of a Network in the Connectivity Format.** (a) A graphical representation of a signaling network's connectivity. (b) The signaling network in (a) written in the *Network Connectivity Format*.

Figure 5: **Examples of marking series and group file formats.** (a) An example marking series dataset in the *comma-separated value* file format. The first row specifies the signaling proteins whose concentrations were measured. Each row thereafter specifies the concentration for a given time step: row i specifies the concentrations for each signaling protein at time step $i - 1$. (b) An example marking group dataset in the *comma-separated value* file format. The first row specifies the signaling proteins whose concentrations were measured. The first column specifies the names for each marking in the group dataset. The numbers in each row specify the concentration measured for each signaling protein in that marking.

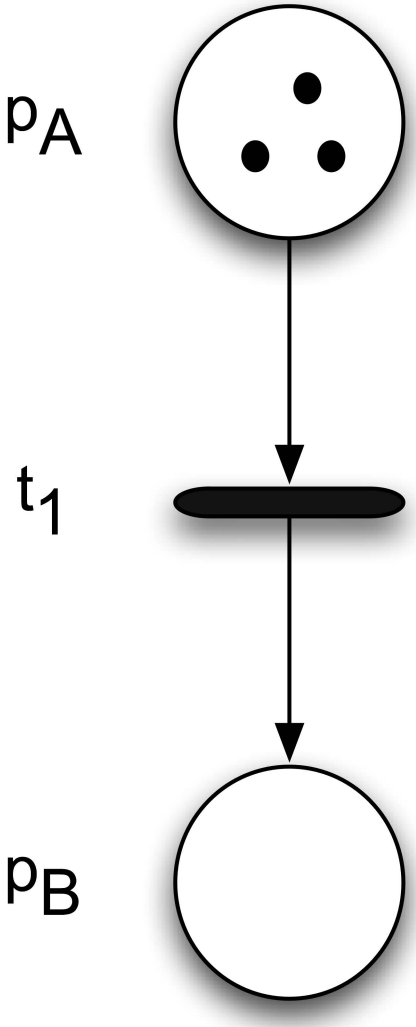
Figure 6: **An example of a Path in the Connectivity Format.** (a) A graphical representation of two signaling paths. (b) The signaling paths in (a) represented in the *Connectivity Format*. Each line corresponds to a single signaling path.

Figure 7: **A comparison of features supported by tools commonly used for signaling network analysis.** The table shows the features and analytical capabilities supported by different tools commonly used for the analysis of signaling networks. Tools included in the comparison are: CellDesigner [20], CellIllustrator [24], CellNetAnalyze [25], COPASI [22], Cytoscape [21], the System Biology Toolkit for Matlab [26], and PathwayOracle.

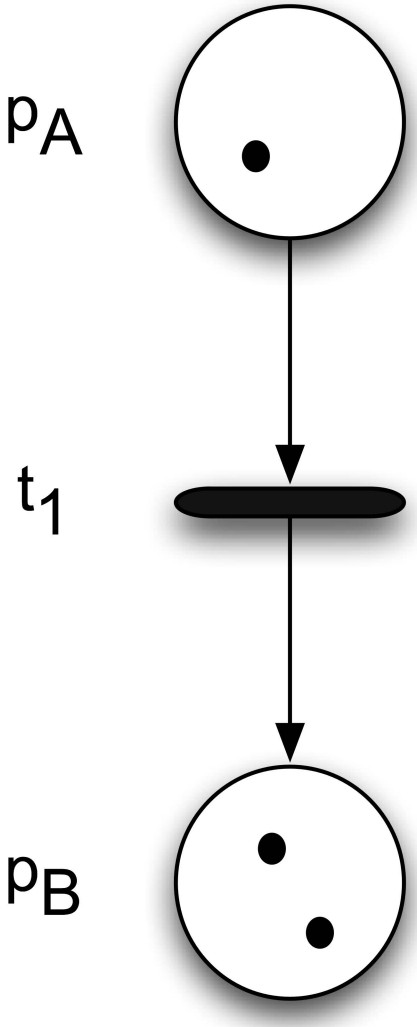
Figure 8: **The tokenized simulator user interface.** (a) The setup window for the tokenized simulator. The simulation is being configured to have two High nodes, EGF and LKB-auto. EGF will be initialized with a token-count of 10, LKB-auto with a token-count of 3. The token-count of AMPK will be zero for the duration of the simulation. (b) The setup window for the differential simulator. Two different scenarios are being compared through simulation: different token assignments are being tried with EGF and LKB-auto, with and without AMPK being fixed low. (c) The plot window for the marking series generated by a simulation. Observe that the signaling nodes whose activity-levels are plotted correspond to those selected in the checklist directly to the left of the plot.

Figure 9: **The path interrogation user interface.** (a) The result window enumerating the set of all paths between *Ras* and *mTOR/raptor*. (b) The main network view showing the selected path highlighted.

Figure 10: **The marking group user interface.** (a) The heat map visualization of a marking group. The selected marking, MDA231-B-DMSO1, is highlighted in blue. (b) The color distribution for the selected marking in the group is applied to the network view in the main window. Note that signaling nodes for which values were not given are not assigned a color on the valid red to green spectrum.



(a)



(b)

Figure 1

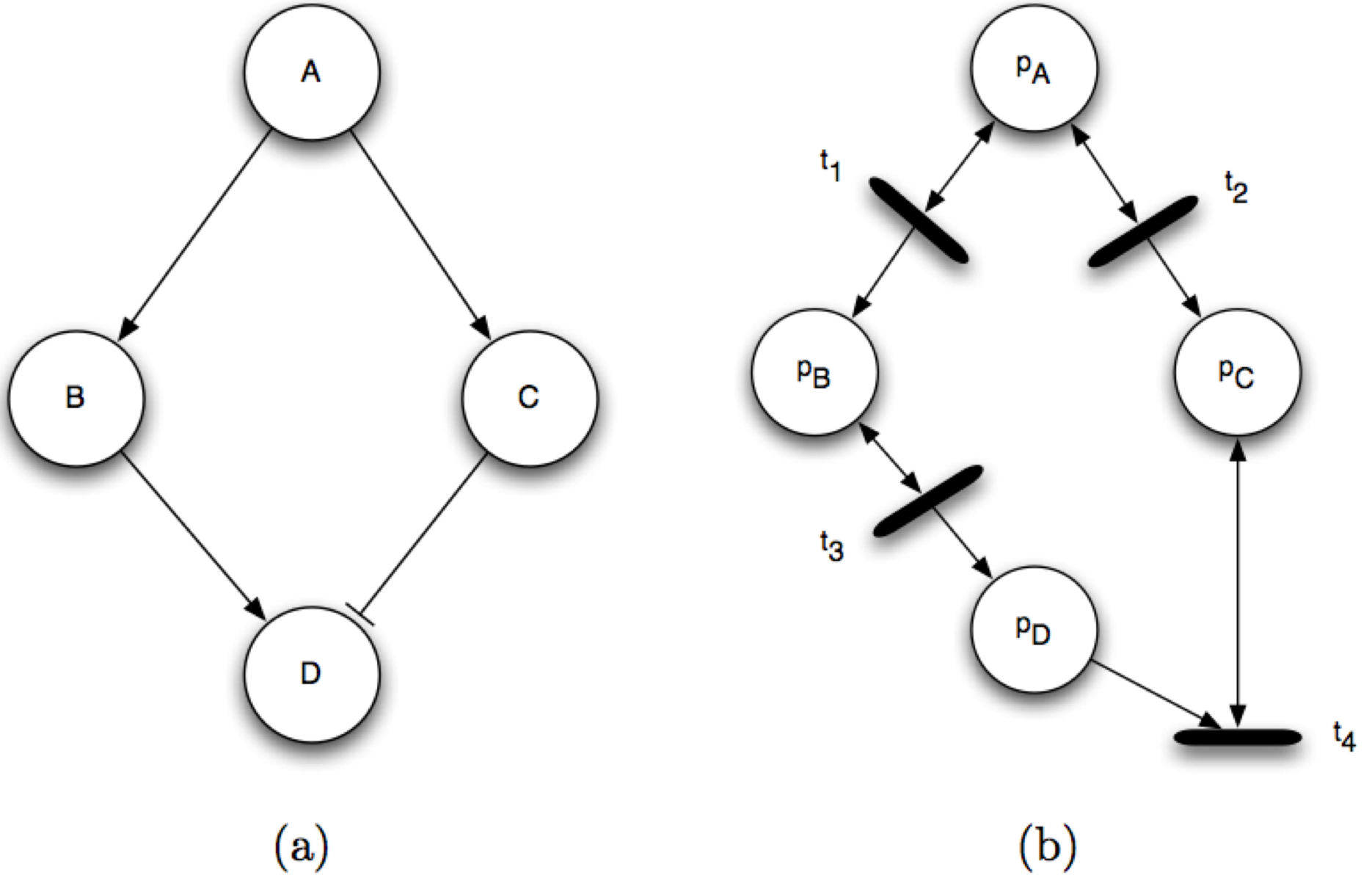
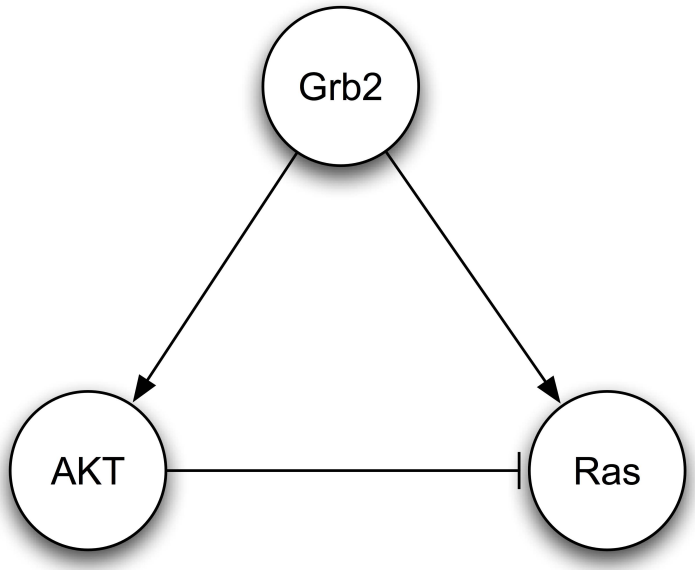
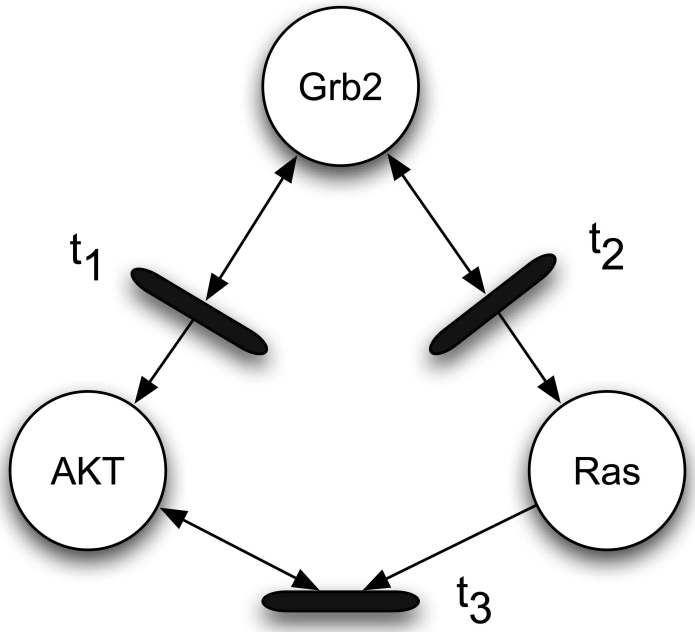


Figure 2



(a)

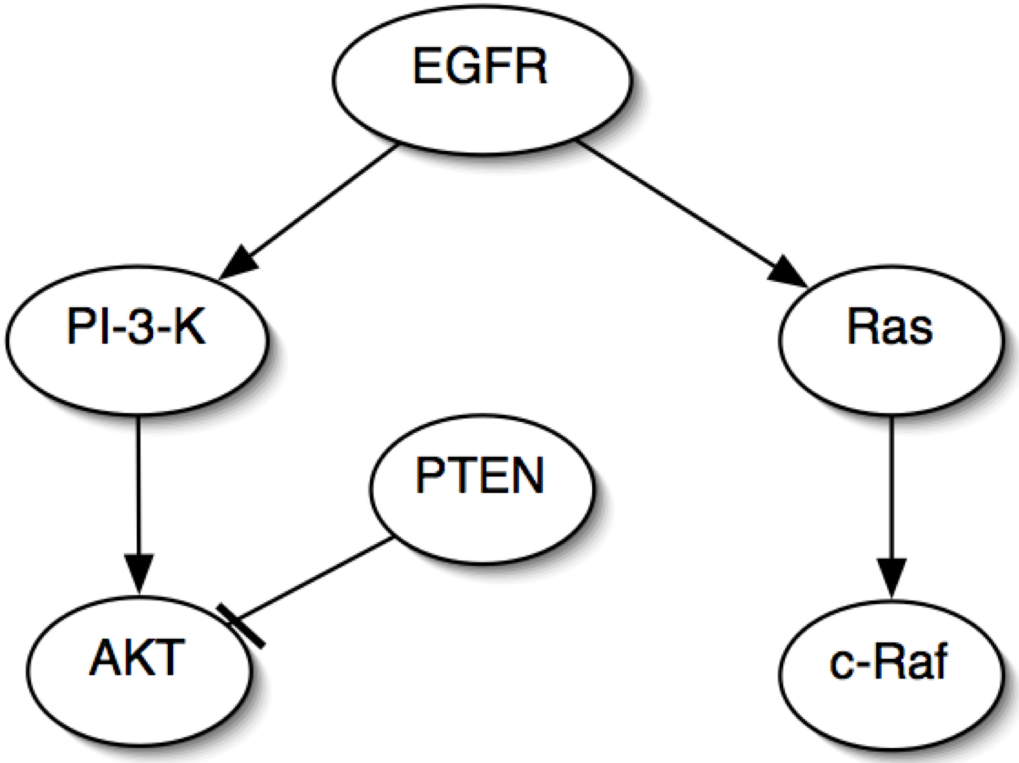


(b)

Proteins	Initial	Block 1			Block 2		
		t_2	t_1	t_3	t_1	t_3	t_2
Grb2	5	5	5	5	5	5	5
AKT	3	3	**6**	6	**10**	10	10
Ras	6	**8**	8	**4**	4	**0**	**2**

(c)

Figure 3



(a)

```
EGFR –> PI-3-K
EGFR –> Ras
PTEN –| PI-3-K
PI-3-K –> AKT
Ras –> c-Raf
```

(b)

Figure 4

mTOR/raptor,	AKT,	EGFR,	RSK
0.3,	0.2,	0.1,	1.1
...	...	...	...
2.1,	0.001,	0.1,	1.5

(a)

,	mTOR/raptor,	AKT,	EGFR,	RSK
DMSO,	0.3,	0.2,	0.1,	1.1
...	...	...	...	...
EGF_30min,	2.1,	0.001,	0.1,	1.5

(b)

Figure 5

(a)

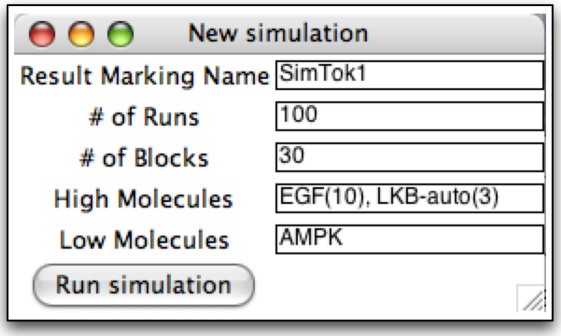
EGFR –> PI-3-K –> AKT –| c-Raf –> MEK
EGFR –> Ras –> c-Raf –> MEK

(b)

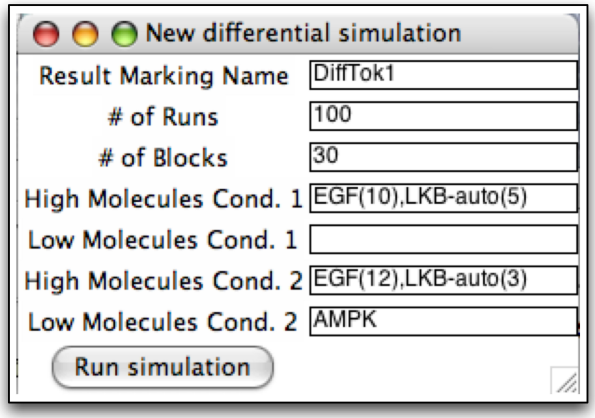
Figure 6

Analysis Type	Features	CellDesigner	CellIllustrator	CellNetAnalyzer	COPASI	Cytoscape	Matlab SB Toolkit	PathwayOracle
	Open Source	-	-	✓	✓	✓	-	✓
	Visual Network Editor	✓	✓	✓	-	✓	-	✓
Experimental Data Analysis	Microarray Visualization	-	-	-	-	✓	-	✓
	Microarray Analysis	-	-	-	-	✓	-	-
Structural Analysis	Structural Statistics	-	-	✓	-	✓	-	-
	Path Finding	-	-	✓	-	✓	-	-
	+/- Path Finding	-	-	✓	-	-	-	✓
	Flux Analysis	-	-	✓	✓	-	-	-
	Boolean Analysis	-	-	✓	-	-	-	-
Dynamic Analysis	ODE Simulation	✓	-	-	✓	-	✓	-
	Hybrid PN Simulation	-	✓	-	-	-	-	-
	Signaling PN Simulation	-	-	-	-	-	-	✓

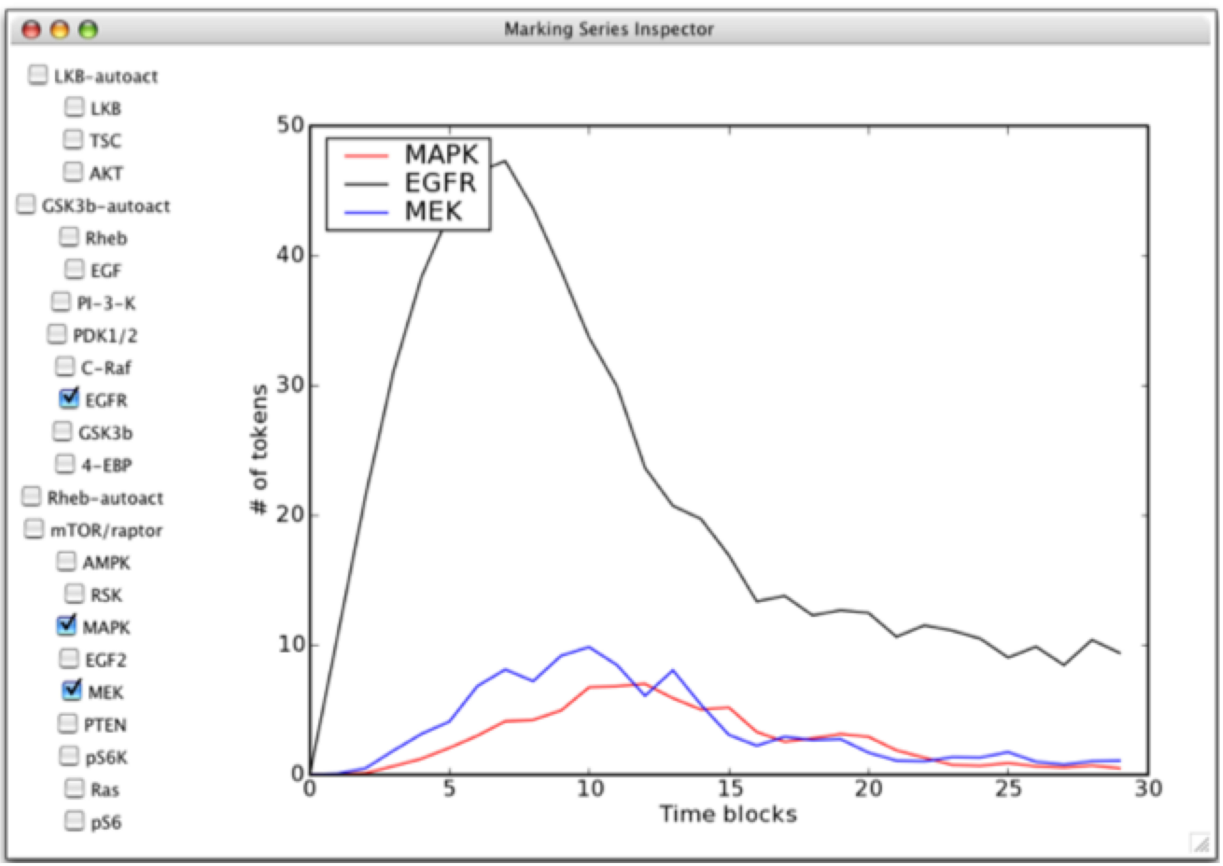
Figure 7


New simulation
Result Marking Name
SimTok1
# of Runs
100
# of Blocks
30
High Molecules
EGF(10), LKB-auto(3)
Low Molecules
AMPK
Run simulation

(a)

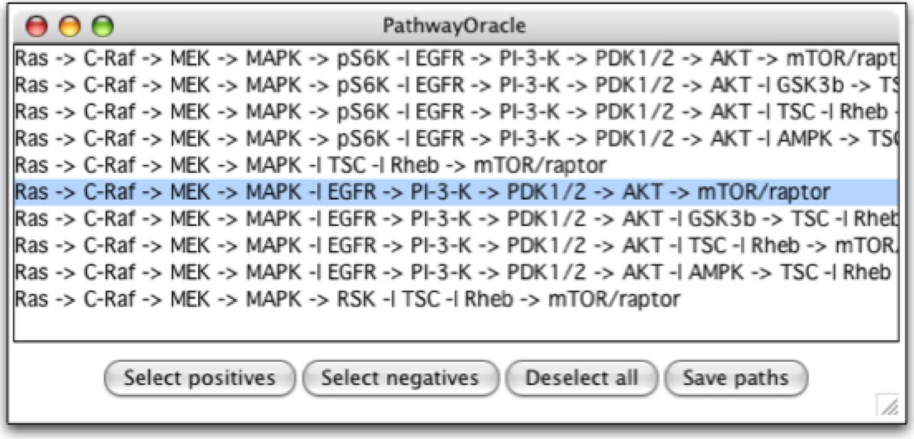

New differential simulation
Result Marking Name
DiffTok1
# of Runs
100
# of Blocks
30
High Molecules Cond. 1
EGF(10),LKB-auto(5)
Low Molecules Cond. 1
High Molecules Cond. 2
EGF(12),LKB-auto(3)
Low Molecules Cond. 2
AMPK
Run simulation

(b)

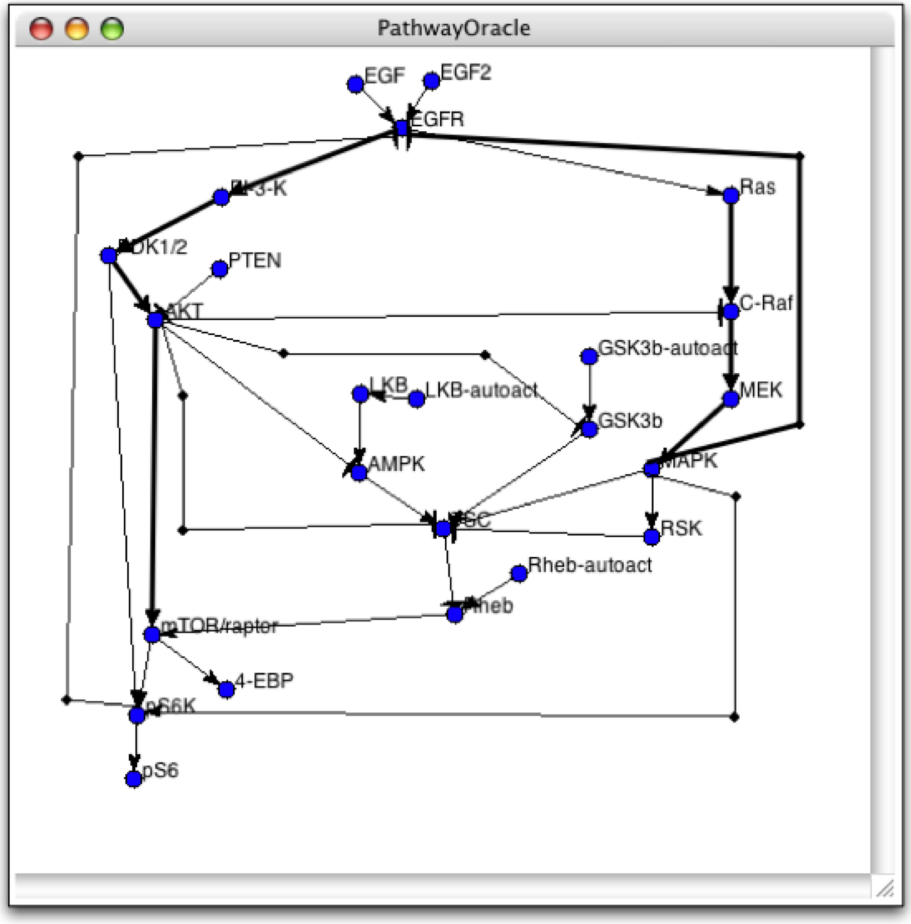

Marking Series Inspector
LKB-autoact
LKB
TSC
AKT
GSK3b-autoact
Rheb
EGF
PI-3-K
PDK1/2
C-Raf
EGFR
GSK3b
4-EBP
Rheb-autoact
mTOR/raptor
AMPK
RSK
MAPK
EGF2
MEK
PTEN
pS6K
Ras
pS6
MAPK
EGFR
MEK
# of tokens
Time blocks
0
10
20
30
40
50
0
5
10
15
20
25
30

(c)

Figure 8

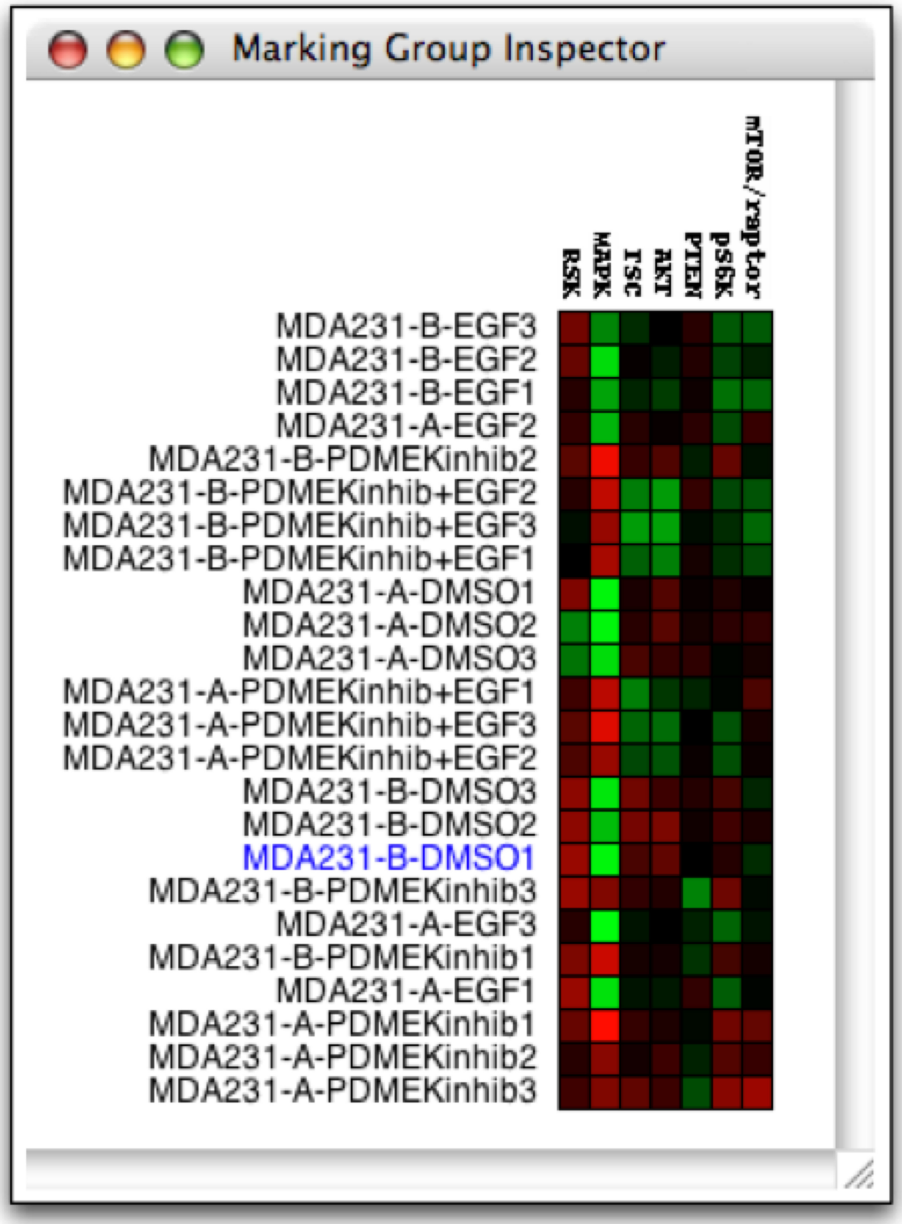

PathwayOracle
Ras -> C-Raf -> MEK -> MAPK -> pS6K -| EGFR -> PI-3-K -> PDK1/2 -> AKT -> mTOR/rapt
Ras -> C-Raf -> MEK -> MAPK -> pS6K -| EGFR -> PI-3-K -> PDK1/2 -> AKT -| GSK3b -> TS
Ras -> C-Raf -> MEK -> MAPK -> pS6K -| EGFR -> PI-3-K -> PDK1/2 -> AKT -| TSC -| Rheb
Ras -> C-Raf -> MEK -> MAPK -> pS6K -| EGFR -> PI-3-K -> PDK1/2 -> AKT -| AMPK -> TSC
Ras -> C-Raf -> MEK -> MAPK -| TSC -| Rheb -> mTOR/raptor
Ras -> C-Raf -> MEK -> MAPK -| EGFR -> PI-3-K -> PDK1/2 -> AKT -> mTOR/raptor
Ras -> C-Raf -> MEK -> MAPK -| EGFR -> PI-3-K -> PDK1/2 -> AKT -| GSK3b -> TSC -| Rheb
Ras -> C-Raf -> MEK -> MAPK -| EGFR -> PI-3-K -> PDK1/2 -> AKT -| TSC -| Rheb -> mTOR
Ras -> C-Raf -> MEK -> MAPK -| EGFR -> PI-3-K -> PDK1/2 -> AKT -| AMPK -> TSC -| Rheb
Ras -> C-Raf -> MEK -> MAPK -> RSK -| TSC -| Rheb -> mTOR/raptor
Select positives
Select negatives
Deselect all
Save paths

(a)

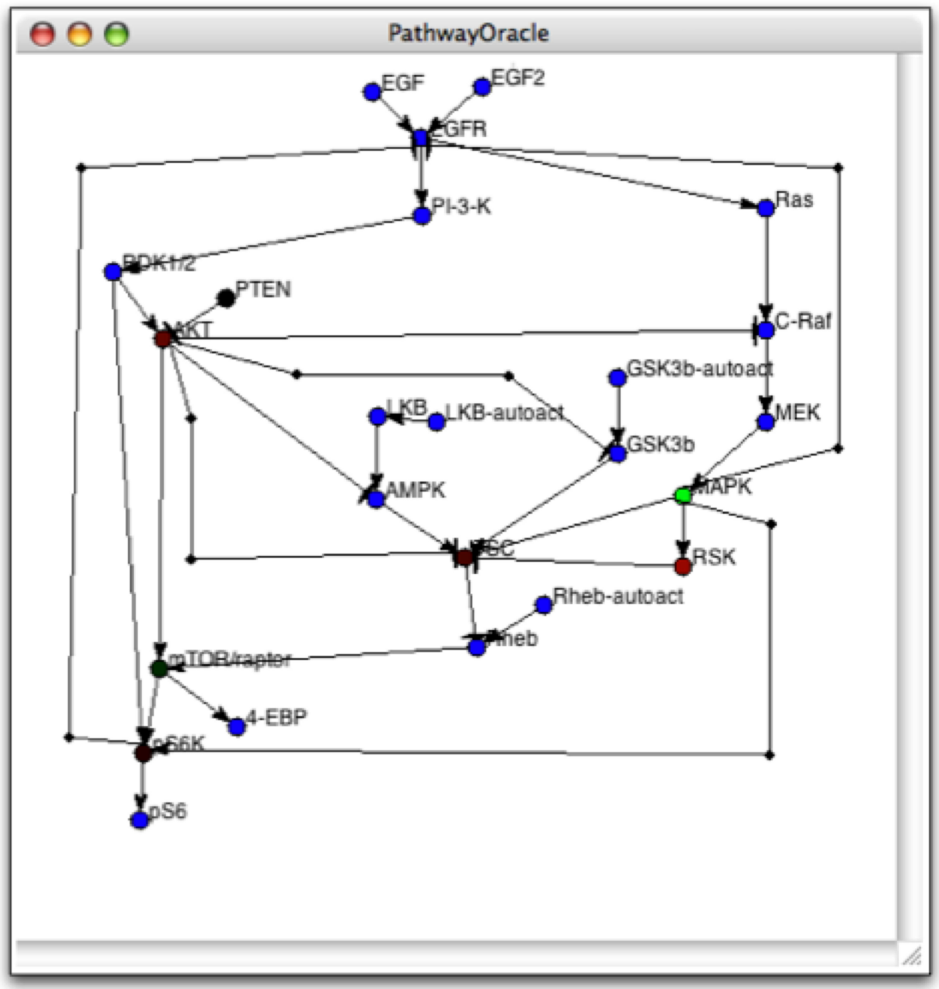

PathwayOracle
EGF
EGF2
EGFR
PI-3-K
Ras
PDK1/2
PTEN
AKT
C-Raf
GSK3b-autoact
LKB
LKB-autoact
MEK
GSK3b
AMPK
MAPK
TSC
RSK
Rheb-autoact
Rheb
mTOR/raptor
4-EBP
pS6K
pS6

(b)

Figure 9


Marking Group Inspector
RSK
MAPK
TSC
AKT
PTEN
pS6K
mTOR/raptor
MDA231-B-EGF3
MDA231-B-EGF2
MDA231-B-EGF1
MDA231-A-EGF2
MDA231-B-PDMEKinhib2
MDA231-B-PDMEKinhib+EGF2
MDA231-B-PDMEKinhib+EGF3
MDA231-B-PDMEKinhib+EGF1
MDA231-A-DMSO1
MDA231-A-DMSO2
MDA231-A-DMSO3
MDA231-A-PDMEKinhib+EGF1
MDA231-A-PDMEKinhib+EGF3
MDA231-A-PDMEKinhib+EGF2
MDA231-B-DMSO3
MDA231-B-DMSO2
MDA231-B-DMSO1
MDA231-B-PDMEKinhib3
MDA231-A-EGF3
MDA231-B-PDMEKinhib1
MDA231-A-EGF1
MDA231-A-PDMEKinhib1
MDA231-A-PDMEKinhib2
MDA231-A-PDMEKinhib3

(a)


PathwayOracle
EGF
EGF2
EGFR
PI-3-K
Ras
PDK1/2
PTEN
AKT
C-Raf
GSK3b-autoact
LKB
LKB-autoact
MEK
GSK3b
AMPK
MAPK
TSC
RSK
Rheb-autoact
Rheb
mTOR/raptor
4-EBP
pS6K
pS6

(b)

Figure 10

Additional files provided with this submission:

Additional file 1: ruthsetal_plos07.pdf, 8652K
http://www.biomedcentral.com/imedia/9980407181769024/supp1.pdf